Journal of Mathematical Psychology 73 (2016) 94–109

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Tutorial

A tutorial on General Recognition Theory

Noah H. Silbert [a,*], Robert X.D. Hawkins [b]

[a] Department of Communication Sciences & Disorders, University of Cincinnati, Cincinnati, OH 45267, USA

[b] Department of Psychology, Stanford University, USA

HIGHLIGHTS

- General recognition theory provides a powerful mathematical framework for analyzing interactions between perceptual dimensions.
- Non-parametric statistical tools allow for analysis of dimensional interactions with minimal assumptions about the underlying perceptual and decisional representations.
- Parametric Gaussian models allow for powerful statistical analyses and compelling model visualizations.

ARTICLE INFO

Article history:
Received 15 September 2015
Received in revised form
27 April 2016

Keywords:
General Recognition Theory
Perceptual separability
Perceptual independence
Tutorial

ABSTRACT

General Recognition Theory (GRT; e.g., Ashby and Townsend, 1986, inter alia) is a two-stage, multidimensional model of encoding and response selection. In this tutorial, we present the basic conceptual and mathematical structure of GRT and review the three notions of dimensional interaction defined in the GRT framework: perceptual independence, perceptual separability, and decisional separability. Experimental protocols and data closely linked to the GRT model are discussed, and two sets of empirical tests of dimensional interaction are presented. These test procedures are illustrated via functions the new R package `mdsdt`.

Contents

* Corresponding author.
E-mail addresses: silbernh@ucmail.uc.edu (N.H. Silbert), rxdh@stanford.edu (R.X.D. Hawkins).

http://dx.doi.org/10.1016/j.jmp.2016.04.011
0022-2496/

1. Introduction

Cognitive processing is fundamentally multidimensional. Even the simplest experimental stimuli must be defined with respect to multiple dimensions. Consider, for example, the auditory perception of a simple tone. Although the amplitude of the tone may be of primary interest to a researcher, a sinusoidal tone must also have frequency, duration, and phase. The perception of amplitude may depend, at least in part, on the frequency, duration, and/or phase of the tone. For example, a 100 Hz tone at a given amplitude will be perceived as substantially more quiet than a 1000 Hz tone with the same amplitude (Plack & Carlyon, 1995). Similarly, at very short times, a longer tone with a given amplitude will be perceived as louder (i.e., as having larger amplitude) than a shorter tone with the same amplitude (Plack & Carlyon, 1995). A full understanding of auditory perception must take this sort of interaction between acoustic dimensions into account. In general, a full understanding of how multiple dimensions are processed in any modality requires an account of the relationships between the dimensions.

General recognition theory (GRT, also known as multidimensional signal detection theory, or MDSDT) is a powerful framework for studying the relationships between dimensions in cognitive processing. It has been used to model perceptual interactions in vision (e.g., Kadlec & Hicks, 1998; Olzak & Wickens, 1997; Thomas, 2001b) and audition (e.g., Silbert, 2012, 2014; Silbert, Townsend, & Lentz, 2009) as well as higher-level processes in memory (e.g., DeCarlo, 2003).

As useful as GRT is, a researcher interested in probing the relationships between dimensions in a particular domain may be intimidated by the mostly technical literature on GRT (Ashby & Townsend, 1986; Kadlec & Townsend, 1992a,b; Silbert, de Jong, Thomas, & Townsend, 2009; Silbert & Thomas, 2013; Thomas, 1995, 1999, 2001a,b). This tutorial seeks to make the powerful tools offered by GRT readily available to the interested researcher. We begin with a brief overview of the structure of GRT. This is followed by a more technical discussion of three different types of dimensional interactions defined in the GRT framework. With these fundamentals in place, the tutorial then takes the reader carefully through multiple procedures for analyzing data and drawing GRT-based inferences based on the results of these analyses. Each procedure is illustrated using corresponding functions in the new R package `mdsdt`, making it fast and practical for the reader to apply these analyses to their own data.

1.1. Basic concepts

GRT is a two-stage model of encoding and response selection. As such, it can be usefully thought of as a multidimensional extension of signal detection theory (SDT; Green & Swets, 1966; Macmillan & Creelman, 2005). GRT, like SDT, models trial-by-trial variation in behavioral responses by assuming a two stage process of noisy encoding followed by deterministic response selection. Although, as noted above, GRT has been used to study non-perceptual processes, it has been applied most frequently to the study of perceptual interactions. We will focus here on perceptual encoding and subsequent response selection, noting here for clarity that this is not intended to imply that GRT is limited to this domain.

We consider noisy perception and deterministic response selection in turn.

1.1.1. Noisy perception

GRT begins with the assumption that a substantial component of perception is stochastic. Although there are multiple possible sources for random perceptual variability (e.g., environmental and/or neural noise; Ashby & Lee, 1993), the specific sources of perceptual variability are not (typically) modeled in GRT.

The basic idea of noisy perception is that from trial to trial, the perceptual effect of a given stimulus varies. Over time, this produces distributions of perceptual effects. With some additional structure (e.g., a mechanism for response selection), the statistical properties of these distributions generate predictions about observed data. This, in turn, allows the researcher to draw inferences about latent perceptual structures from observed data. Properties of these latent perceptual structures then license inferences about the presence or absence of perceptual dimensional interactions.

In the most general form of GRT, no specific assumptions are made about the properties of the stochastic component of perception. However, it is common to assume that noisy perception is usefully modeled with multivariate Gaussian distributions, analogous to the univariate Gaussian distributions of SDT. We discuss statistical analyses of GRT data for both cases below.

1.1.2. Deterministic response selection

Noisy perception alone does not allow a perceptual model to make contact with data. We also need a mechanism for selecting a response given a noisy perceptual effect. In GRT, this mechanism typically takes the form of decision bounds, or curves that partition perceptual space into mutually exclusive response regions.

Response selection is deterministic in GRT. When a (random) perceptual effect occurs in a particular response region, the response associated with that region is always selected. Response probabilities are given by the joint properties of the (noisy) perceptual distributions and (deterministic) decision bounds.

More technically, the probability of emitting response X to stimulus Y is given by the integral of the perceptual distribution corresponding to stimulus Y over the response region for response X.

1.1.3. Gaussian GRT model visualization

In order to illustrate noisy perception and deterministic response selection, we provide a visual depiction of a 2×2 Gaussian GRT model. By '2×2', we mean a model (and associated stimulus set) with four perceptual distributions comprising the factorial combination of two levels on each of two dimensions. For example, we may be interested in using GRT to analyze possible

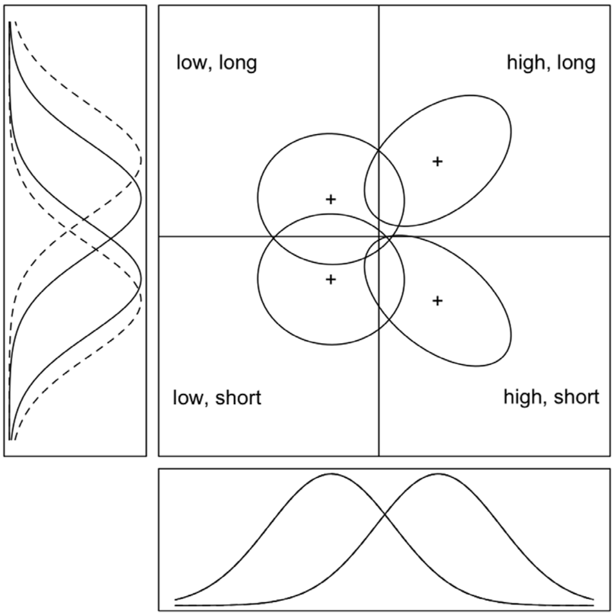

Fig. 1. Illustrative example of a 2 × 2 Gaussian GRT model. In the main panel, the *x*-axis is the modeled perceptual dimension corresponding to frequency, the *y*-axis is the modeled perceptual dimension corresponding to duration, the plus-signs indicate the means of the perceptual distributions, and circles/ellipses indicate the distributions' equal-likelihood contours. The vertical and horizontal lines indicate decision bounds. Labels indicate response regions. In the smaller panels to the left and on the bottom, the marginal perceptual distributions are plotted, with the marginal distributions at the 'low' and 'high' levels on each dimension plotted with solid and dotted lines, respectively. See text for additional details.

interactions between frequency and duration in the perception of sound. Silbert, Townsend et al. (2009) probed the perception of these dimensions in broadband noise in their first experiment. In this case, listeners simultaneously identified the frequency range and duration of the stimuli, and the stimulus set consisted of the factorial combination of *low* and *high* frequency levels (490–1490 Hz and 510–1510 Hz, respectively) and *short* and *long* duration levels (250 ms and 300 ms, respectively). Hence, the full set had four stimuli: *low*, *short*; *low*, *long*; *high*, *short*; and *high*, *long*.

Taking a top-down view, we can get a fairly comprehensive view of the Gaussian GRT model through plots of decision bounds and equal likelihood contours for each of the perceptual distributions.[1] Fig. 1 illustrates a possible 2 × 2 model. In the large square panel, the plus-signs indicate the means of the perceptual distributions, while the circles and ellipses are equal likelihood contours, or sets of points on each perceptual distribution at the same height above the plane. The shape of the equal likelihood contours indicate presence or absence of correlation in each distribution, with circles indicating zero correlation and ellipses indicating non-zero correlation. The dimensions along which the perceptual distributions and decision bounds are defined are modeled perceptual dimensions corresponding to the physical dimensions of the stimuli.

When taking such a top-down view of a GRT model, it is important to keep in mind that random perceptual effects for a particular stimulus may occur inside or outside the equal likelihood contour, and that perceptual effects inside the contour are more likely than those outside (i.e., the perceptual distribution's peak is inside the contour, at the mean of the distribution, and the height of the bell curve describing the distribution decreases with distance from the mean). A single equal likelihood contour provides a convenient visualization of a bivariate perceptual distribution, but it is a simplification.

The vertical and horizontal lines indicate decision bounds, and the labels indicate the response regions. In the smaller, rectangular panels below and to the left of the square panel, the marginal perceptual distributions are illustrated, with the solid lines corresponding to the closer distributions and the dashed lines corresponding to the more distant distributions (e.g., in the left panel, the solid lines illustrate the marginal *low*, *short* and *low*, *long* distributions, while the dashed lines illustrate the marginal *high*, *short* and *high*, *long* distributions).

As described above, the probability of a particular response to a given stimulus is given by the integral of the appropriate perceptual distribution in the appropriate response region. So, for example, the probability of a 'low, long' response to the *high*, *long* stimulus is the double integral of the top-right distribution over the top-left response region. Hence, the configuration of four perceptual distributions and two decision bounds produces 16 predicted response probabilities, one for each response region for each perceptual distribution.

This top-down view of a GRT model also provides a convenient visualization of dimensional interactions, which we consider in the following section.

1.2. Dimensional interactions

As noted above, GRT provides powerful tools for analyzing perceptual interactions. The assumptions of GRT – noisy perception and deterministic response selection – provide the basis for three logically distinct ways in which dimensions may or may not interact in multidimensional perception. In this section, we review these notions of (lack of) interaction between dimensions defined in the GRT framework: *perceptual independence*, *perceptual separability*, and *decisional separability*. We emphasize that these three notions are theoretical constructs, and so are not directly observable. The goal in GRT-based analyses is to draw inferences about these unobservable constructs from properties of appropriately collected data. The primary aim of this tutorial is to illustrate for the reader how to do this.

1.2.1. Perceptual independence

Perceptual independence holds between two dimensions within a given perceptual distribution if, and only if, the perceptual effects on each dimension are stochastically independent within that perceptual distribution. If frequency and duration are perceptually independent in *low*, *short* stimuli, for example, perception of higher or lower frequency is not statistically associated with perception of shorter or longer duration when *low*, *short* stimuli are presented. If perceptual independence fails for frequency and duration in *low*, *short* stimuli, on the other hand, then there is such a statistical relationship when *low*, *short* stimuli are presented.

In Fig. 1, perceptual independence holds in the (circular) A_1B_1 (*low*, *short*) and A_1B_2 (*low*, *long*) distributions, while it fails in the (elliptical) A_2B_1 (*high*, *short*) and A_2B_2 (*high*, *long*) distributions.

Mathematically, let the joint density of perceptual effects for stimulus A_iB_j be $f_{A_iB_j}(x, y)$, where A_i indicates the ith level of the stimulus on the x dimension and B_j indicates the jth level on the y dimension. In the example discussed above, A indicates frequency and B indicates duration, with A_1 = low frequency, A_2 = high

[1] In the 2 × 2 model, the means and variances of the perceptual distributions are not both identifiable. Any change in means can be offset with an appropriate scaling of variances, and vice versa. We briefly return to this issue below. In order to fix the scale of the 2 × 2 model, we set all marginal variances equal to one.

frequency, B_1 = short duration, and B_2 = long duration. The marginal densities on each dimension are then:

$$\int_{-\infty}^{\infty} f_{A_iB_j}(x, y)dy = g_{A_iB_j}(x) \quad (1)$$

$$\int_{-\infty}^{\infty} f_{A_iB_j}(x, y)dx = g_{A_iB_j}(y).$$

Perceptual independence holds if, and only if Eq. (2) holds for all values of x and y:

$$f_{A_iB_j}(x, y) = g_{A_iB_j}(x)g_{A_iB_j}(y). \quad (2)$$

Because perceptual independence is a property of a single, unobserved perceptual distribution, it can, in principle, hold or fail for each perceptual distribution regardless of whether it holds or fails in any other.

1.2.2. Perceptual separability

Unlike perceptual independence, perceptual separability defines relationships *across* perceptual distributions. Perceptual separability holds on one dimension with respect to another dimension if the marginal perceptual distributions on the former do not vary across levels of the latter. So, for example, if the perception of frequency does not vary across levels of duration (i.e., if the perception of frequency is equivalent in short and long stimuli), we would say that frequency is perceptually separable from duration.

In Fig. 1, perceptual separability holds on the x dimension with respect to the y dimension (i.e., frequency is perceptually separable from duration). The A_1B_1 and A_1B_2 distributions are aligned vertically, as are the A_2B_1 and A_2B_2 distributions, and the marginal distributions illustrated in the bottom panel are clearly coincident across levels of duration (B_1 and B_2).

On the other hand, perceptual separability fails on the y dimension with respect to the x dimension (i.e., duration is *not* perceptually separable from frequency). The marginal perceptual distributions of perceptual effects on the y dimension are closer together at the A_1 level (low frequency; solid lines in the left panel) than they are at the A_2 level (high frequency; dashed lines in the left panel), indicating that the perceptual salience of duration varies across levels of frequency (A_1 and A_2).

Mathematically, perceptual separability holds on the x dimension with respect to the y dimension (for A with respect to B) if, and only if, Eq. (3) holds for all values of x and y, while perceptual separability holds on the y dimension with respect to the x dimension (for B with respect to A) if, and only if, Eq. (4) holds for all values of x and y:

$$g_{A_iB_1}(x) = g_{A_iB_2}(x) \qquad i = 1, 2 \quad (3)$$

$$g_{A_1B_j}(y) = g_{A_2B_j}(y) \qquad j = 1, 2. \quad (4)$$

Eq. (3) states that the marginal perceptual effects at each level of the x dimension (i.e., A_1 and A_2) are identically distributed at each level of the y dimension (i.e., B_1 and B_2). Eq. (4) states that the marginal perceptual effects at each level of the y dimension (i.e., B_1 and B_2) are identically distributed at each level of the x dimension (i.e., A_1 and A_2).

Because perceptual separability concerns relationships between marginal perceptual distributions on a given dimension, it may hold or fail on either dimension regardless of whether or not perceptual independence holds in any of the distributions.

1.2.3. Decisional separability

Finally, in the GRT framework, dimensions may interact with respect to response selection. Decisional separability holds on one dimension with respect to another if, and only if, the decision bound partitioning the former is parallel to the coordinate axis on the latter. Both decision bounds in Fig. 1 exhibit decisional separability.

If decisional separability holds, then decisions on one dimension do not depend in any way on another dimension. Complicating matters, there are a number of ways in which decisions on one dimension could depend on another dimension. A model could have linear bounds that are not parallel to any of the coordinate axes. In this case, decisions on one dimension would depend on multiple dimensions of the perceptual space. On the other hand, decisions on one dimension could depend on *decisions* on another dimension (e.g., responses concerning the frequency of a sound may depend on whether or not a sound has been labeled 'short' or 'long'), in which case the model would have piecewise (linear) decision bounds (see, e.g., Silbert & Thomas, 2013).[2]

Another possibility is that responses are selected based on the relative likelihood of a given perceptual effect having been produced by each possible stimulus (i.e., by a Bayesian response rule). In this case, decisions are not directly determined by decision bounds. This type of model may be described as having *implicit* piecewise (quadratic or linear) decision bounds (see, e.g., Silbert & Thomas, 2014).

Recent theoretical work in the GRT framework has shown that failures of decisional separability are not, in general, identifiable (Silbert & Thomas, 2013; Thomas & Silbert, 2014).[3] However, under the assumption that decisional separability holds, failures of perceptual independence and perceptual separability *are* identifiable. In addition, many failures of perceptual independence and/or perceptual separability are identifiable regardless of whether or not decisional separability holds. In particular, failures of perceptual separability that are caused by changes in the salience of one dimension across levels of the other are identifiable whether or not decisional separability holds. This type of failure of perceptual separability is illustrated in Fig. 1, wherein the distinction between short and long sounds is more salient at the high frequency level than at the low frequency level. Real-world examples of this kind of failure of perceptual separability has been found with speech stimuli, as well (Silbert, 2012, 2014). For detailed discussion of these issues, see Silbert and Thomas (2013). In this tutorial, we will assume that decisional separability holds and focus on methods for analyzing perceptual interactions (or lack thereof).

2. How to collect and analyze data with GRT

In order to draw GRT-based inferences about possible interactions between cognitive dimensions, we need appropriate data, and we need appropriate analytic techniques. In this section, we begin by describing the basic factorial identification experimental paradigm. We then discuss in some detail what the data collected in this kind of experiment looks like. We then give a brief

[2] If decisional separability fails on just one dimension in this way, the piecewise bounds will be vertical or horizontal, depending on which dimension exhibits the failure. If decisional separability fails on *both* dimensions in this way, there are problematic ambiguities with respect to how the discontinuous decision bounds should be connected. See Silbert and Thomas (2013) for a detailed discussion and illustration of this issue.

[3] There have been recent claims that a multi-level extension of GRT can identify failures of decisional separability (Soto & Ashby, 2015; Soto, Vucovich, Musgrave, & Ashby, 2104), but the proposed model is mathematically equivalent to a model in which decisional separability holds (Silbert & Thomas, submitted for publication).

Table 1
Observed identification-confusion frequencies for Subject 3 from Experiment 1 of Silbert, Townsend et al. (2009).

Stimulus	Response			
	a_1b_1	a_1b_2	a_2b_1	a_2b_2
A_1B_1	159	33	46	12
A_1B_2	20	186	5	39
A_2B_1	21	9	191	29
A_2B_2	3	22	22	203

Table 2
Observed identification-confusion frequencies for Subject 7 from Experiment 2 of Silbert, Townsend et al. (2009).

Stimulus	Response			
	a_1b_1	a_1b_2	a_2b_1	a_2b_2
A_1B_1	186	22	41	1
A_1B_2	16	180	26	28
A_2B_1	42	32	149	27
A_2B_2	1	28	13	208

overview of three approaches to analyzing factorial identification data in order to draw inferences about perceptual interactions: non-parametric statistical tests, comparison of marginal signal detection theory sensitivity and bias parameters, and Gaussian GRT model parameter estimation and model comparison. Before explaining these three approaches in detail, it is important to ensure that the analyst is familiar with the structure of factorial identification experiments and data.

2.1. Experimental paradigms

The nature of the perceptual distributions and decision bounds in a GRT model must be closely matched to data collected in an associated experimental protocol. The most common approach to using GRT to draw inferences about interactions between perceptual dimensions is the 'feature-complete factorial' or 'complete identification' paradigm, mentioned briefly above with respect to the 2×2 model and stimulus specification.

The feature-complete factorial paradigm employs stimuli and responses defined by the factorial combination of N levels on each of M dimensions. In the simplest (and most common) instantiation of this, the stimuli and responses are defined with respect to two levels on each of two dimensions. We will focus primarily on this case, though we will also briefly discuss 3×3 factorial identification and experiments and data in which there are more response levels than stimulus levels on each dimension. Although we do not discuss it here, the basic ideas developed for analyzing simple factorial models and data also generalize to cases in which there are two or more levels on a larger number of dimensions (e.g., Kadlec & Townsend, 1992a,b).

Although there are other approaches that fall within the broad framework of GRT, including a substantial body of literature on two-choice categorization using 'decision bound theory' (e.g., Ashby & Gott, 1988; Ashby & Maddox, 1992; Maddox & Ashby, 1993), this tutorial will focus on factorial identification tasks, data, and associated statistical testing and modeling techniques.

2.2. Factorial identification data

As described above, stimuli and responses in the 2×2 factorial identification paradigm are defined with respect to two levels on each of two dimensions. As above, we denote stimulus levels on the first dimension with A_i, $i = 1, 2$ and on the second dimension with B_j, $j = 1, 2$, and we denote responses on the first and second dimension with lowercase a_i and b_j, respectively. Hence, the stimulus consisting of the first level on the x dimension and the second level on the y dimension would be denoted A_1B_2, while a response indicating the first level on each dimension would be denoted a_1b_1, and so on.

The identification-confusion responses collected in the 2×2 task are cross-tabulated to indicate the number of times that each response was given to presentations of each stimulus. A 2×2 GRT confusion matrix is depicted in Table 1.

This confusion matrix is from Subject 3 of Experiment 1 from Silbert, Townsend et al. (2009). In this experiment, listeners identified broadband noise stimuli consisting of the factorial combination of two levels of frequency range (A_1: 490–1490 Hz, A_2: 510–1510 Hz) and two levels of duration (B_1: 250 ms, B_2: 300 ms). Henceforth, we will refer to this experiment as F $\times$ D (Frequency $\times$ Duration).

We also consider the confusion matrix from Subject 7 of Experiment 2 from Silbert, Townsend et al. (2009), depicted in Table 2, in which listeners identified 13-component harmonic stimuli consisting of the factorial combination of two levels of fundamental frequency (A_1: 150 Hz, A_2: 152 Hz) and the location of a spectral prominence (B_1: 850 Hz, B_2: 1050 Hz). We will refer to this experiment as P $\times$ T (Pitch $\times$ Timbre). These two data sets are included in the `mdsdt` package as `silbert09a` and `silbert09b`, respectively.

In each confusion matrix, the rows correspond to stimuli and the columns to responses. Hence, the number in the ith row and jth columns indicates the number of times that response j was given to stimulus i. In the matrices presented above, the number of a_2b_1 responses given to the A_2B_2 stimulus – a total of 22 in the first matrix and 13 in the second – is given in the 4th row and 3rd column of each matrix.

2.3. Drawing GRT inferences

There are a number of approaches one may take to using factorial identification data to draw inferences about the presence or absence of interactions between perceptual dimensions (i.e., the presence or absence of perceptual independence and/or perceptual separability). One option is to use tests of *marginal response invariance* and *report independence*[4] to investigate the presence or absence of perceptual interactions (e.g., Ashby & Townsend, 1986; Silbert, 2010; Thomas, 2001a,b). While this approach requires only very weak assumptions about the nature of the underlying perceptual distributions, it is relatively opaque with respect to the nature of any perceptual interactions, requiring a number of logical steps to connect observed (estimated) statistics to unobserved perceptual structures.

On the other hand, one may fit a model, or fit and compare multiple models, by estimating the parameters of perceptual distributions and decision bounds directly (e.g., Silbert, 2012, 2014; Thomas, 2001a,b; Wickens, 1992). This approach requires one to make strong assumptions about the functional form of the underlying perceptual distributions and decision bounds, but it also allows the researcher to clearly illustrate the modeled perceptual structures and use powerful statistical tests of hypothesized interactions. Model fitting requires specialized software (e.g., Silbert, 2012, 2014; Wickens, 1992); GRT model parameters cannot, in general, be estimated directly from the count data in a confusion matrix.

In between these two extremes, one may use marginal signal detection analyses. These require fairly strong assumptions about

4 In a number of earlier studies in the GRT framework, this was called *sampling independence*. Following Townsend, Houpt, and Silbert (2012), we use the term *report independence*, since the estimated statistic relates directly to observed reports and only indirectly to unobserved perceptual sampling.

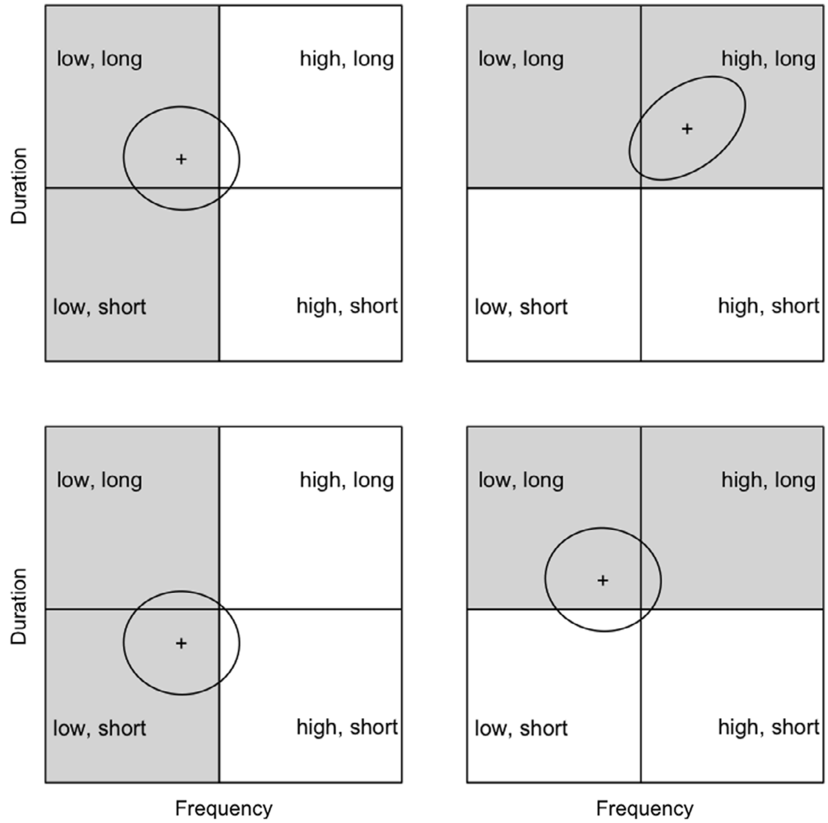

Fig. 2. Visualization of the marginal response probabilities used in the definition of marginal response invariance, with the F × D stimulus and response labels (i.e., A_1 = low frequency, A_2 = high frequency, B_1 = short duration, B_2 = long duration). The shaded regions in the two panels on the left illustrate the marginal response probabilities $p(a_1|A_1B_1)$ (bottom) and $p(a_1|A_1B_2)$ (top). The shaded regions in the two panels on the right illustrate the marginal response probabilities $p(b_2|A_1B_2)$ (bottom) and $p(b_2|A_2B_2)$ (top).

the nature of the underlying perceptual structures, though they may be carried out on statistics estimated directly from confusion matrices. Although marginal signal detection analyses have a sound mathematical basis (e.g., Kadlec & Townsend, 1992a), recent work suggests that these analyses are, on the one hand, largely redundant with respect to analyses of marginal response invariance and report independence, and, on the other hand, prone to problematically high rates of false alarms (Silbert & Thomas, 2013). Consequently, we do not discuss marginal signal detection analyses in detail in this tutorial, and tools for conducting marginal signal detection analyses are *not* included in the `mdsdt` package.

In this tutorial, we focus on marginal response invariance, report independence, parameter estimation, and model comparison, all of which are conveniently implemented in the `mdsdt` package. In Section 3, we define marginal response invariance and report independence, we present statistical tests of each, and we illustrate the implementation of these tests via functions defined in the `mdsdt` package. In Section 4, we discuss Gaussian GRT model parameter estimation and model comparison, illustrating each of these procedures with functions from the `mdsdt` package.

3. Marginal response invariance and report independence

Unlike perceptual independence, perceptual separability, and decisional separability, marginal response invariance and report independence are directly observable. Each is defined with respect to relationships between particular subsets of identification data. In this section, we begin by defining marginal response invariance and report independence. We then describe two closely related statistical approaches to testing for failures of perceptual separability and perceptual independence using these definitions. Finally, we illustrate the application of these tests to the data presented in the two confusion matrices above using the `mdsdt` package.

3.1. Definitions

Marginal response invariance is typically defined with respect to correct marginal responses on each level of each dimension. Marginal response invariance holds for a given level on a given dimension if the marginal probability of identifying that level on that dimension is the same across levels of the other dimension. Put more concretely, in the F × D data given in Table 1, marginal response invariance holds for *low frequency* if the probability of responding 'low' (i.e., a_1, in conjunction with either 'short' – b_1 or 'long' – b_2) is the same for the *low*, *short* and *low*, *long* stimuli (i.e., across the levels of the duration dimension).

Mathematically, marginal response invariance holds for the ith level on the first dimension if the following equation holds:

$$\begin{aligned} p(a_i|A_iB_1) &= p(a_ib_1|A_iB_1) + p(a_ib_2|A_iB_1) \\ &= p(a_ib_1|A_iB_2) + p(a_ib_2|A_iB_2) = p(a_i|A_iB_2). \end{aligned} \quad (5)$$

Analogously, marginal response invariance holds for the jth level on the second dimension if the following holds:

$$\begin{aligned} p(b_j|A_1B_j) &= p(a_1b_j|A_1B_j) + p(a_2b_j|A_1B_j) \\ &= p(a_1b_j|A_2B_j) + p(a_2b_j|A_2B_j) = p(b_j|A_2B_j). \end{aligned} \quad (6)$$

The marginal response probabilities used in the definition of marginal response invariance for levels a_1 (*low frequency*) and b_2 (*long duration*) are illustrated in Fig. 2. In each panel, the shaded region indicates a region over which a perceptual distribution is

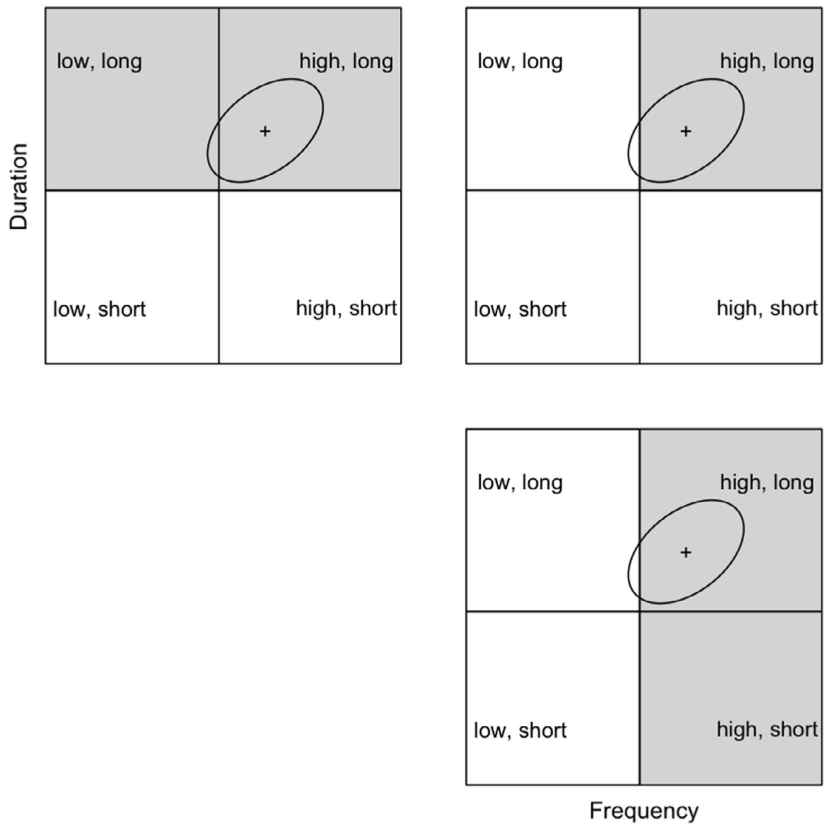

Fig. 3. Visualization of the marginal and joint response probabilities used in the definition of report independence, with the F × D stimulus and response labels (i.e., A_1 = low frequency, A_2 = high frequency, B_1 = short duration, B_2 = long duration). The shaded regions in the top-left and bottom-right illustrate the marginal response probabilities $p(b_2|A_2B_2)$ and $p(a_2|A_2B_2)$, respectively. The shaded region in the top-right panel illustrates the joint response probabilities $p(a_2b_2|A_2B_2)$.

integrated, with the appropriate perceptual distribution indicated by a single equal likelihood contour. The left two panels illustrate the marginal response probabilities involved in the definition of marginal response invariance for level a_1, while the right two panels illustrate the marginal response probabilities for level b_2.

On a given dimension, marginal response invariance may hold or fail at each level whether or not it holds at the other level. So, for example, the marginal probability of correctly identifying the *low frequency* level (a_1) may be invariant across levels of duration, while the marginal probability of correctly identifying the *high frequency* level on the first dimension (a_2) may vary across levels of duration.

Marginal response invariance provides information about the presence or absence of perceptual separability. If decisional separability and perceptual separability both hold, together they imply that marginal response invariance will hold (Ashby & Townsend, 1986). Hence, under the assumption that decisional separability holds, marginal response invariance holds at both levels of a given dimension if and only if perceptual separability holds on that dimension. Thus, we may use marginal response invariance as a critical test for the hypothesis of perceptual separability: if we find that marginal response invariance fails at either level of a given dimension, we can conclude that perceptual separability fails on that dimension.

Report independence provides information about the presence or absence of perceptual independence, which may hold or fail for each stimulus. Hence, report independence is defined with respect to a single stimulus. Report independence holds for a given stimulus if the probability of correctly identifying that stimulus is equal to the product of the marginal probabilities of accurately identifying the level of each component of that stimulus. For example, in the F × D data given in Table 1, report independence holds for the *high frequency*, *long duration* stimulus (i.e., A_2B_2) if the probability of correctly identifying this stimulus (i.e., the probability of correctly identifying both components of this stimulus) is equal to the product of the marginal probabilities of correctly identifying *low frequency*, on the one hand, and *long duration*, on the other.

Mathematically, report independence holds for stimulus A_iB_j if the following equation holds:

$$\begin{aligned} p(a_ib_j|A_iB_j) &= p(a_i|A_iB_j) \times p(b_j|A_iB_j) \\ &= \Big[p(a_ib_1|A_iB_j) + p(a_ib_2|A_iB_j)\Big] \\ &\quad \times \Big[p(a_1b_j|A_iB_j) + p(a_2b_j|A_iB_j)\Big]. \end{aligned} \tag{7}$$

The marginal and joint response probabilities used in the definition of report independence for a_2 (*high frequency*), b_2 (*long duration*), and a_2b_2 (*high frequency*, *long duration*) are illustrated in Fig. 3. In each panel, the shaded region indicates a region over which a perceptual distribution is integrated, with the relevant perceptual distribution indicated by an equal likelihood contour, in this case for stimulus A_2B_2. The top-left panel illustrates the marginal response probability at level b_2, the bottom-right panel illustrates the marginal response probability for level a_2, and the top-right panel illustrates the joint response probability for a_2b_2.

If both perceptual independence and decisional separability hold, this implies that report independence will hold, as well (Ashby & Townsend, 1986). Hence, under the assumption that decisional separability holds, the presence or absence of report independence is a direct indicator of the presence or absence of perceptual independence.

3.2. Statistical analysis of marginal response invariance and report independence

Of course, the equations defining marginal response invariance and report independence are unlikely to hold exactly in any given data set, even if perceptual separability and perceptual independence are present. This is because the probabilities estimated from subsets of the data in a confusion matrix are unlikely to be exactly equal to the probabilities determined by a particular configuration of underlying perceptual distributions and decision bounds. Hence, we need a statistical treatment of these theoretically derived properties of the data. The R Package `mdsdt` provides functions that carry out statistical tests of marginal response invariance and report independence. We discuss the statistical foundation of each of these tests and illustrate their application using the `mdsdt` package in turn.

3.2.1. Estimation of response probabilities

We denote the cell in the ith row and jth column of a confusion matrix $M_{i,j}$. The count data given in the confusion matrices above can be used to estimate various response probabilities by calculating appropriate response proportions. Such estimated probabilities are used to calculate the test statistics for marginal response invariance and report independence, as described below. We provide a brief review of basic response probability estimation here.

The estimated probability of response a_1b_2 given that stimulus A_1B_1 was presented is denoted $\hat{p}(a_1b_2|A_1B_1)$. This is estimated as the number of a_1b_2 responses to the A_1B_1 stimulus divided by the total number of presentations of stimulus A_1B_1, which is the sum across all responses in the row for stimulus A_1B_1. Since stimulus A_1B_1 is in row 1 and response a_1b_2 is in the column 2 in the matrices given above, this is calculated as follows:

$$\hat{p}(a_1b_2|A_1B_1) = \frac{M_{1,2}}{\sum_{j=1}^{4} M_{1,j}} = \frac{M_{1,2}}{M_{1,\cdot}}. \tag{8}$$

For the first matrix presented above, this estimated probability is $\frac{33}{250} = 0.132$, while for the second matrix it is $\frac{22}{250} = 0.088$. Other response probabilities are estimated similarly, with appropriate adjustments to the response count and row sum appearing the numerator and denominator, respectively.

We will also make use of marginal response probabilities. For example, we might need to estimate the probability of responding a_1 on the first dimension, regardless of the response on the second dimension, given the presentation of stimulus A_1B_2 (i.e., the probability of correctly identifying the level of the first dimension of this stimulus). We denote this $\hat{p}(a_1|A_1B_2)$. Since stimulus A_1B_2 is in row 2 and the two relevant responses a_1b_1 and a_1b_2 are in columns 1 and 2 in the matrices above, this is calculated as follows:

$$\begin{aligned}\hat{p}(a_1|A_1B_2) &= \hat{p}(a_1b_1|A_1B_2) + \hat{p}(a_1b_2|A_1B_2)\\ &= \frac{M_{2,1}}{M_{2,\cdot}} + \frac{M_{2,2}}{M_{2,\cdot}} = \frac{M_{2,1}+M_{2,2}}{M_{2,\cdot}}.\end{aligned} \tag{9}$$

In the first matrix given above, this estimated probability is $\frac{20+186}{250} = \frac{206}{250} = 0.824$, while in the second it is $\frac{16+180}{250} = \frac{196}{250} = 0.784$.

Similarly, the probability of correctly identifying the level of the second dimension of this stimulus would be calculated as follows:

$$\begin{aligned}\hat{p}(b_2|A_1B_2) &= \hat{p}(a_1b_2|A_1B_2) + \hat{p}(a_2b_2|A_1B_2)\\ &= \frac{M_{2,2}}{M_{2,\cdot}} + \frac{M_{2,4}}{M_{2,\cdot}} = \frac{M_{2,2}+M_{2,4}}{M_{2,\cdot}}.\end{aligned} \tag{10}$$

In the first matrix given above, this estimated probability is $\frac{186+39}{250} = \frac{225}{250} = 0.9$, while in the second it is $\frac{180+28}{250} = \frac{208}{250} = 0.832$.

Table 3
z test results for marginal response invariance given the F × D data (Table 1) and P × T data (Table 2).

Experiment		Response level			
		A_1	A_2	B_1	B_2
F × D	z	−1.55	−0.72	−0.84	0.00
	p	0.12	0.48	0.40	1.00
P × T	z	1.36	−4.98	4.35	−3.97
	p	0.17	<0.001	<0.001	<0.001

3.2.2. A statistical test of marginal response invariance

We can test marginal response invariance at the ith level of the first dimension via the normalized difference of estimated marginal response probabilities (Thomas, 2001a):

$$z_{a_i} = \frac{\hat{p}(a_i|A_iB_1) - \hat{p}(a_i|A_iB_2)}{\sqrt{\hat{p}_{i*}\left(1-\hat{p}_{i*}\right)\left(N_{A_iB_1}^{-1} + N_{A_iB_2}^{-1}\right)}}. \tag{11}$$

Here, the estimated marginal response probabilities are as defined above, $N_{A_iB_j}$ is the number of presentations of stimulus A_iB_j, and $\hat{p}_{i*}$ is the overall proportion of a_i responses to the A_iB_1 and A_iB_2 stimuli.

We can test marginal response invariance at the jth level of the second dimension analogously:

$$z_{b_j} = \frac{\hat{p}(b_j|A_1B_j) - \hat{p}(b_j|A_2B_j)}{\sqrt{\hat{p}_{j*}\left(1-\hat{p}_{j*}\right)\left(N_{A_1B_j}^{-1} + N_{A_2B_j}^{-1}\right)}}. \tag{12}$$

All terms are defined analogously, with $\hat{p}_{j*}$ indicating the overall proportion of b_j responses to the A_1B_j and A_2B_j stimuli.

This z test statistic approximates a standard normal random variate, so a standard normal table will provide critical test statistic and p values. In the `mdsdt` package, the function `mriTest()` provides the z statistics and p values for tests of marginal response invariance at each level of each dimension. Note that this function (and all others) in the `mdsdt` package assume that the confusion matrix is organized as the two matrices given above (i.e., with the A_1, B_1 stimulus and response in the first row and column, respectively, the A_1, B_2 stimulus and response in the second row and column, etc...):

```
> data(silbert09a)
> mriTest(silbert09a)
  response      z p.value
1   (A1,-) -1.554   0.120
2   (A2,-) -0.715   0.475
3   (-,B1) -0.841   0.400
4   (-,B2)  0.000   1.000
```

As can be seen in Table 3, marginal response invariance holds for each level of each dimension for the F × D data, whereas three of the four tests fail for the P × T data, indicating a failure of perceptual separability on both dimensions. The marginal response invariance test only needs to fail at one level of a dimension (e.g. at level A_2 in the P × T experiment) to provide evidence for an overall failure of perceptual separability on that dimension.

The z statistic for the test of marginal response invariance is a signed statistic. In Eqs. (11) and (12), the marginal response proportion at the *second* level of a given dimension is subtracted from the marginal response proportion at the *first* level. Hence, a negative z statistic indicates that the latter is smaller than the former, and a positive z statistic indicates the opposite.

More concretely, in the P × T data, the z statistic for the A_2 response level is negative, indicating that the marginal proportion

Table 4
Observed and marginal frequencies for stimulus A_1B_1 (*low*, *short*) from Table 1.

Frequency range	Duration		
	b_1 = short	b_2 = long	
a_1 = low	159	33	192
a_2 = high	46	12	58
	205	45	250

Table 5
Expected frequencies for stimulus A_1B_1 (*low*, *short*) from Table 1.

Frequency range	Duration	
	b_1 = short	b_2 = long
a_1 = low	157.44	34.56
a_2 = high	47.56	10.44

of A_2 ('low fundamental frequency') responses at the B_1 level (low spectral prominence) is less than the marginal proportion of A_2 responses at the B_2 level (high spectral prominence). From this, we can infer that perceptual separability fails here because the A_2, B_2 perceptual distribution is shifted farther toward the A_2 ('high fundamental frequency') response level than is the A_2, B_1 perceptual distribution.

The z statistic for the b_1 ('low spectral prominence') response level is positive, indicating that the marginal proportion of b_1 responses is greater at the A_1 level (low fundamental frequency) than at the A_2 level (high fundamental frequency). We can infer from this that the A_1, B_1 (low, low) perceptual distribution is shifted farther toward the b_1 ('low prominence') response level than is the A_2, B_1 (high fundamental, low prominence) perceptual distribution. This is also clearly visible in Fig. 8.

3.2.3. A statistical test of report independence

We can test report independence by carrying out a χ^2 test of independence on the row of data corresponding to a stimulus of interest. For stimulus A_iB_j, we can reorganize the data from the appropriate row of the full confusion matrix to construct a 2 × 2 contingency table. For example, we can reorganize the data from the A_1B_1 stimulus (i.e., the 1st row) from the first matrix given above:

The χ^2 test statistic is defined with respect to e_{ij} and $M_{i,j}$, the expected and observed response frequencies, respectively (Thomas, 2001a):

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(M_{ij} - e_{ij}\right)^2}{e_{ij}}. \tag{13}$$

The expected response frequency is defined under the assumption that report independence holds:

$$e_{ij} = N_{A_iB_j} \frac{M_{i,\cdot}}{N_{A_iB_j}} \frac{M_{\cdot,j}}{N_{A_iB_j}} = \frac{M_{i,\cdot} M_{\cdot,j}}{N_{A_iB_j}}. \tag{14}$$

Using this formula, the expected frequencies for Table 4 are:

Using the observed and expected frequencies from Tables 4 and 5, respectively, and the formula from Eq. (13), the $\chi^2(1)$ statistic for stimulus A_1B_1 from Table 1 is 0.37, $p = 0.54$.[5] A non-significant test result here indicates that we cannot reject the hypothesis that report independence holds (i.e., the data are at least roughly consistent with report independence). As noted above, if decisional separability also holds (or if it assumed to hold), report independence is a direct indicator of perceptual independence.

[5] There is one degree of freedom for this test because there are two rows ($r = 2$) and two columns ($c = 2$) in the reorganized tables of observed and expected frequencies, and $df = (r - 1)(c - 1)$.

Table 6
χ^2 statistics from report independence tests run on the F × D data (Table 1) data and the P × T data (Table 2).

Experiment		Stimulus			
		A_1B_1	A_1B_2	A_2B_1	A_2B_2
F × D	χ^2	0.37	0.11	5.79	0.12
	p	0.54	0.74	0.02	0.73
P × T	χ^2	2.81	48.42	22.50	0.29
	p	0.09	<0.001	<0.001	0.60

Report independence tests are performed in the R package using the `riTest` function, which computes χ^2 statistics for each stimulus in the given confusion matrix:

```
> data(silbert09a)
> riTest(silbert09a)
  stimulus chi.2 p.value
1  (A1,B1) 0.370   0.543
2  (A1,B2) 0.110   0.740
3  (A2,B1) 5.793   0.016
4  (A2,B2) 0.123   0.725
```

The χ^2 statistics and p values for each stimulus in Tables 1 and 2 are summarized in Table 6.

For the F × D experiment, the data for three of the four stimuli are consistent with report independence, exhibiting extremely small χ^2 values. By way of contrast, the A_2B_1 (*high*, *short*) stimulus has a moderately large (and statistically significant, with $\alpha = 0.05$) χ^2 value.

On the other hand, for the P × T experiment, the data for only two stimuli are consistent with report independence (A_1B_1 and A_2B_2), while the data for the other two produce very large, and so statistically significant, χ^2 values, indicating failure of perceptual independence for those stimuli.

4. Gaussian GRT parameter estimation and model comparison

The tests of marginal response invariance and report independence described above require only weak assumptions about the nature of the noisy perception and deterministic response selection underlying a confusion matrix, but they provide relatively indirect indications of the presence or absence of perceptual separability and perceptual independence. If we are willing to make stronger assumptions about the underlying GRT model, we can clearly visualize the full set of relationships between perceptual dimensions and draw upon well-established model comparison tools for testing various hypotheses about the presence or absence of perceptual interactions. The standard set of stronger assumptions employed in GRT model fitting are that the perceptual distributions are bivariate Gaussian distributions and that the decision bounds are linear. These additional assumptions constrain the researcher to express the presence or absence of perceptual independence and perceptual separability through particular values of and relationships between to Gaussian probability density function parameters. We employ these assumptions here.

As discussed above, the signs and magnitudes of the z statistics for testing marginal response invariance can be used to infer particular arrangements of perceptual distributions. On the other hand, the χ^2 statistic for testing report independence does not provide information beyond a binary indication of the presence or absence of perceptual independence (and decisional separability), though, when report independence fails, the estimated probabilities used to define report independence provide information regarding the nature of the underlying failure of perceptual independence.

In addition to enabling powerful statistical tests of the presence or absence of perceptual independence and perceptual separability, Gaussian GRT model fitting and comparison allow for much more clear and concise visualization of the precise configuration of modeled perceptual distributions. GRT model fitting and comparison also comprise the most common approach to using GRT to analyze identification-confusion data.

4.1. Parameter estimation

In the basic 2×2 Gaussian model, there are 12 parameters to estimate.[6] We must estimate a two-dimensional mean $\boldsymbol{\mu}_{a_i b_j}$ and a correlation $\rho_{a_i b_j}$ for each stimulus. Because the model is translation, rotation, and scale invariant,[7] and because we are interested in the relationships between the dimensions with respect to the locations and shapes of the perceptual distributions, we set the decision bounds at $x = y = 0$ and estimate the means of the perceptual distributions relative to these axes.[8] As noted above, because the means and marginal variances are not simultaneously identifiable in the 2×2 case, we also fix all marginal variances at unity.

Assumptions of perceptual separability and independence can be directly encoded as constraints on the estimated parameters (e.g., equal means on the x dimension across levels of the y dimension; correlations equal to 0), yielding models with fewer free parameters (and therefore less model complexity), but a poorer fit to the data in general. Statistical model comparison can then be used to decide whether or not the assumptions embodied by any particular set of constraints are justified.

In order to fit a Gaussian GRT model using the `mdsdt` package in R, we simply call the `fit.grt()` function with a confusion matrix as the first (and possibly sole) input argument. If the confusion matrix is the only input argument, this function estimates all possible mean and correlation parameters. Additional arguments may be provided to fit constrained variants of the model with assumptions of perceptual separability and/or independence enforced. In particular, we can use `PS_x = TRUE` to fit a model with perceptual separability on the x dimension (i.e., invariant x means across levels of y) and `PS_y = TRUE` for the model with perceptual separability on the y dimension (i.e., invariant y means across levels of x). We have included three options for perceptual independence: (1) `PI = 'none'`, the default option, where we make no assumptions about independence, (2) `PI = 'same.rho'`, where a single correlation parameter is estimated for all four perceptual distributions (i.e., perceptual independence fails in the *same way* for all stimuli), and (3) `PI = 'all'`, where we assume that perceptual independence holds for all stimuli.

It is important to note that the input confusion matrix to the `fit.grt()` function should be arranged as are the two matrices given above (i.e., with the first row corresponding to the A_1B_1, bottom-left perceptual distribution in Fig. 1, the second row corresponding to the A_1B_2, top-left distribution, the third row to the A_2B_1, bottom-right distribution, and the fourth row corresponding to the A_2B_2, top-right distribution).

Given a particular set of assumptions (e.g., perceptual separability on the x dimension and perceptual independence in all four distributions), the `fit.grt()` function returns a `grt` object containing parameter estimates, standard errors of the parameter estimates, and various measures of the goodness of fit of the model to the data (e.g., AIC, AIC_c, BIC; Burnham & Anderson, 2002). Individual components of the object may be obtained using the `$` notation. So, for example, if the fitted model object is called `fitted.model`, a matrix containing the estimates of the means, standard deviations, and correlations for each distribution can be extracted by typing `fitted.model$dists`. Observed response counts, predicted response probabilities, the parameter gradient, and the inverse of the parameter covariance matrix may be obtained by typing `fitted.model$fit`. AIC, AIC_c, and BIC fit statistics may be obtained from the fitted model object in analogous manner (i.e., via `fitted.model$AIC`, `fitted.model$AIC.c`, and `fitted.model$BIC`). Fit statistics may also be obtained by calling, e.g., `GOF(fitted.model, ''AIC'')`. A summary of the estimates, standard errors, and fit statistics is given by `summary(fitted.model)`.

Example R sessions fitting and comparing four models to the data given in Tables 1 and 2 are shown in Figs. 5 and 6, respectively.

4.2. Model comparison

The procedures for testing (failures of) perceptual separability and perceptual independence with Gaussian model fits employ somewhat different reasoning than do the tests of marginal response invariance and report independence described above. The z and χ^2 tests of marginal response invariance and report independence yield test statistics and associated p-values which license the *rejection* of the null hypotheses that separability and/or independence hold. In Gaussian GRT model comparison, by way of contrast, the default (unconstrained) model allows for *failures* of both separability and independence, whereas alternative, constrained models force separability and/or independence to hold. The goodness of fit of a constrained model is, by necessity, no better than (i.e., as good or worse than) the goodness of fit of more flexible model. However, any reduction in goodness of fit due to the imposition of constraints on a model may be (statistically) justified. If a constrained alternative model turns out to be justified, then we can conclude that separability and/or independence hold.

There are two approaches to testing perceptual separability and perceptual independence using model comparisons. On the one hand, we may use likelihood ratio tests to compare *nested* models. On the other hand, we can use fit statistics that balance goodness of fit and model complexity to compare non-nested models.

4.2.1. Likelihood ratio tests

A model is nested with respect to another model if the former is created by constraining a subset of the latter's free parameters. For example, a 2×2 Gaussian GRT model with no constraints on the mean or correlation parameters may be constrained to force perceptual separability to hold by making the means of the four perceptual distributions form a rectangle, i.e., by making the following equalities hold[9]:

$$\begin{aligned} \mu_{x,A_1B_1} &= \mu_{x,A_1B_2} \\ \mu_{x,A_2B_1} &= \mu_{x,A_2B_2} \\ \mu_{y,A_1B_1} &= \mu_{y,A_2B_1} \\ \mu_{y,A_1B_2} &= \mu_{y,A_2B_2}. \end{aligned} \tag{15}$$

6 Note that, in a 4×4 confusion matrix, there are 12 degrees of freedom. Hence, the GRT model with 12 free parameters is a saturated model, and so should, if reasonable parameter values can be found, predict the response probabilities perfectly. Because the model is saturated, it cannot be tested against a more general baseline model, though it can serve as a baseline model against which to test more constrained models, as discussed below.

7 The model produces the same predicted response probabilities if it is, in its entirety, shifted by an arbitrary amount δ, rotated by an arbitrary degree θ, or scaled by an arbitrary factor α.

8 Another possibility would be to fix one distribution (e.g., A_1B_1) at the origin and estimate the location of the decision bounds rather than the mean vector for this distribution.

9 The first two equations correspond to the `fit.grt()` argument `PS_x=T`, while the second two equations correspond to the argument `PS_y=T`.

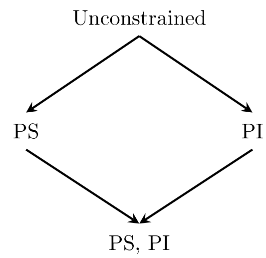

Fig. 4. Illustration of nesting relationships between the unconstrained model, the model with perceptual separability (PS) but not perceptual independence (PI), the model with perceptual independence but not perceptual separability, and the model with both perceptual separability and perceptual independence. Arrows point from general to nested, constrained models.

In this case, the model with perceptual separability is nested with respect to the unconstrained model. The unconstrained model may be similarly constrained to force perceptual independence to hold, i.e., by making $\rho_{A_1B_1} = \rho_{A_1B_2} = \rho_{A_2B_1} = \rho_{A_2B_2} = 0$.[10] Finally, the model with (just) perceptual separability may be constrained to force perceptual independence to hold, and the model with (just) perceptual independence may be constrained to force perceptual separability to hold, creating two more nested model relationships. These relationships are illustrated in Fig. 4.

In each case, the `anova()` function may be used to test whether or not the imposed constraints worsen the model fit sufficiently to reject the constrained model. The `anova()` function takes the fitted *constrained* model object as the *first* argument and the more general (*less constrained*) fitted model object as the *second* argument, and it returns a table with each model's G^2 statistics, the difference of the two model's G^2 statistics (which serves as the test statistic), the difference in degrees of freedom of the two models (which serves as the degrees of freedom for the test statistic), and the p value. A statistically significant test statistic indicates that the more constrained model worsens the model fit 'too much', i.e., that the constraints are *not* justified and that the more general, less constrained model provides a better account of the data. A non-statistically significant test statistic indicates that the imposed constraints are justified and that the flexibility of the more general model is not necessary in order to account for the data. Hence, given a non-statistically significant test, the constrained model can be taken to provide a statistically reasonable account of the data, and the psychological substance of the model's constraints may be thereby inferred (e.g., the presence of perceptual separability).

For each of the two data sets given above, R sessions involving the four likelihood ratio tests comparing a full (i.e., unconstrained) model, a model with (just) perceptual separability, a model with (just) perceptual independence, and a model with both perceptual separability and perceptual independence are illustrated in Figs. 5 and 6.

4.2.2. Goodness of fit and model complexity

As noted above, we can also directly compare non-nested models. The `mdsdt` package includes three fit statistics that provide a measure of the goodness of fit of the model to the data balanced against a penalty for model complexity. In increasing order of the size of the penalty term (i.e., the degree to which the fit statistic penalizes model complexity), these statistics are defined as follows, with $\mathcal{L}$ indicating the log likelihood value for the maximum likelihood parameter estimates, K indicating the number of free parameters in the model, and n indicating the number of observations in the data:

$$\begin{aligned} AIC &= -2\mathcal{L} + 2K \qquad (16)\\ AIC_c &= AIC + \frac{2K(K+1)}{n-K-1} \\ BIC &= -2\mathcal{L} + \log(n)K. \end{aligned}$$

Burnham and Anderson (2002) provide a detailed presentation and discussion of the relative merits of these three fit statistics. In the illustrations given here, and in the authors' experience using these fit statistics for Gaussian GRT model comparisons, all three tend strongly to lead to the same inferences. For each of these fit statistics, the model with the smallest statistic is the best.

4.2.3. Interpretation of model comparison results

The R sessions shown in Figs. 5 and 6 illustrate the extraction of *AIC* statistics from the fitted model objects. In Fig. 5, we see that the full (default) model has an AIC value of 1708, the model with (just) perceptual separability on each dimension has an AIC of 1704, the model with (just) perceptual independence in each distribution has an AIC of 1707, and the model with both perceptual separability and perceptual independence has an AIC of 1703. The AIC_c and BIC statistics exhibit the same pattern. Hence, the model with both perceptual separability and perceptual independence provides the best balance between goodness of fit and model complexity.

On the other hand, in Fig. 6, the full (default) model has an AIC value of 1758, the model with (just) perceptual separability on each dimension has an AIC of 1804, the model with (just) perceptual independence in each distribution has an AIC of 1814, and the model with both perceptual separability and perceptual independence has an AIC of 1866. The AIC_c and BIC statistics exhibit the same pattern. Hence, in this example, the full model provides the best balance of goodness of fit and model complexity.

The best fitting model from these two R sessions are illustrated in Figs. 7 and 8, organized as was the illustrative example given in Fig. 1. Note that, because perceptual separability holds in the model illustrated in Fig. 7, the marginal distributions on each dimension are equal across levels of the other dimension, so no dashed lines are visible.

The data analyzed in the R session illustrated in Fig. 5 are from the F × D data, given in Table 1. The model comparison results indicate that both perceptual separability and perceptual independence hold. From the fact that perceptual separability holds we conclude that the perception of frequency range does not depend on the duration of a noise, and that the perception of the duration of a noise does not depend on the frequency range spanned by that noise. And from the fact that perceptual independence holds we conclude that the perception of frequency range and duration are uncorrelated within each of the four stimuli. Note that these results are consistent with the results of the tests of marginal response invariance and report independence given above in Tables 3 and 6, respectively (with the exception of the one statistically significant test of report independence). The best fitting Gaussian model predicts, correctly, that marginal response invariance and report independence should hold for this data set.

The data being analyzed in the R session illustrated in Fig. 6 are from the P × T data, given in Table 2. In this case, both perceptual separability and perceptual independence fail. The failure of perceptual separability indicates that perception of pitch (fundamental frequency) depends on the level of the spectral prominence dimension and that perception of timbre (spectral

[10] This constraint corresponds to the `fit.grt()` argument `PI=''all''`. The argument `PI=''one.rho''` would allow the four correlation parameters to equal zero or any non-zero value between −1 and 1.

```
> library(mdsdt)
# Import data
> data(silbert09a)
# Fit unconstrained (full) model
> full.mod.a = fit.grt(silbert0a)
# Fit an alternative model assuming (just) perceptual separability
> PS.mod.a = fit.grt(silbert09a, PS_x = T, PS_y = T)
# Fit an alternative model assuming (just) perceptual independence
> PI.mod.a = fit.grt(silbert09a, PI = 'all')
# Fit an alternative model assuming perceptual separability
# and perceptual independence
> PS.PI.mod.a = fit.grt(silbert09a, PS_x = T, PS_y = T, PI = 'all')
# conduct likelihood ratio tests for nested model pairs
# testing (just) perceptual separability vs. no constraints
> anova(PS.mod.a, full.mod.a)
              G2         df DG2    df p-val
PS.mod.a      11255.408  4
full.mod.a    11251.863  0  3.545  4  0.4711
# testing (just) perceptual independence vs. no constraints
> anova(PI.mod.a, full.mod.a)
              G2         df DG2    df p-val
PI.mod.a      11258.838  4
full.mod.a    11251.863  0  6.974  4  0.1373
# testing PS and PI vs. just PS
> anova(PS.PI.mod.a,PS.mod.a)
                G2         df DG2    df p-val
PS.PI.mod.a     11262.34   8
PS.mod.a        11255.408  4  6.932  4  0.1395
# testing PS and PI vs. just PI
> anova(PS.PI.mod.a,PI.mod.a)
                G2         df DG2    df p-val
PS.PI.mod.a     11262.34   8
PI.mod.a        11258.838  4  3.502  4  0.4776
# Alternatively, check AIC statistics for each model
> c(full.mod.a$AIC, PS.mod.a$AIC, PI.mod.a$AIC, PS.PI.mod.a$AIC)
AIC     AIC     AIC     AIC
1708.1  1703.7  1707.1  1702.6
# Visualize the best fitting model (PS.PI.mod.a)
> plot(PS.PI.mod.a, marginals=TRUE)
```

Fig. 5. Example analysis session in R, using Gaussian GRT fitting procedures and the `silbert09a` data provided in the `mdsdt` package. None of the nested models are statistically significantly worse fitting than their more general counterparts, and the model assuming perceptual separability and perceptual independence has the lowest AIC value (as well as the lowest AIC_c and BIC values), all of which indicates that this model provides the best overall account of the data. Hence, we conclude that perceptual separability and perceptual independence held for this participant in this task. The last line produces the fitted model plot shown in Fig. 7.

prominence location) depends on the level of the fundamental frequency dimension. The *low*, *high* (f0, spectral prominence) stimulus is perceived as having higher f0 than the *low*, *low* stimulus and lower spectral prominence than the *high*, *high* stimulus. Analogously, the *high*, *low* stimulus is perceived as having lower f0 than the *high*, *high* stimulus and higher spectral prominence than the *low*, *low* stimulus. The failure of perceptual independence is most strongly exhibited in the *low*, *high* and *high*, *low* perceptual distributions, both of which have large magnitude negative correlations, indicating that, within each of these stimuli, lower-valued random perceptual effects on one dimension tended to be accompanied by higher-valued random perceptual effects on the other dimension.

The combined effect of these failures of perceptual separability and independence can be interpreted as a partial perceptual conflation of these two dimensions. When the two dimensions 'matched' – in the *low*, *low* and *high*, *high* stimuli – the dimensions interacted synergistically, producing highly salient perceptual effects. When the two dimensions 'mismatched' – in the *low*, *high* and *high*, *low* stimuli – the two dimensions were in conflict, producing highly confusable perceptual effects.

These results are also consistent with the corresponding tests of marginal response invariance and report independence described above in Tables 3 and 6, respectively. Marginal response invariance fails in three of the four tests (holding at level A_1), with the signs of the z statistics consistent with the configuration of perceptual distributions illustrated in Fig. 8. Similarly, report independence fails dramatically for the *low*, *high* (A_1B_2) and *high*, *low* (A_2B_1) stimuli.

5. Other paradigms

Although the focus of this tutorial has been analysis of data from the 2 × 2 factorial identification paradigm, the GRT framework may be modified so that it can be used to analyze data from other experimental paradigms, as well. Some of the mathematical theory has been elaborated for GRT models with more than two dimensions (e.g., Kadlec & Townsend, 1992b), and Gaussian GRT models have been fit to 3 × 3 identification data (e.g., Ashby & Lee, 1991; Thomas, Altieri, Silbert, Wenger, & Wessels, 2015). The `fit.grt()` function in the `mdsdt` package can be used to estimate parameters for $M \times M$ factorial identification data, and there are two 3 × 3 data sets included in the package. These data sets are from the two observers discussed in Thomas et al. (2015). They can be loaded by calling `data(''thomas15a'')` and `data(''thomas15b'')`.

The `thomas15a` and `thomas15b` data consist of identification counts for a 3 × 3 array of faces varying on nose length and eye width. When the data are loaded, these names will refer to `xtabs` tables with the three nose length and eye width levels defining

```
# Import data (assumes library(mdsdt) has already been invoked)
> data(silbert09b)
# Fit unconstrained model
> full.mod.b = fit.grt(silbert09b)
# Fit an alternative model assuming perceptual separability
> PS.mod.b = fit.grt(silbert09a, PS_x = T, PS_y = T)
# Fit a model assuming perceptual independence
# in all four distributions
> PI.mod.b = fit.grt(silbert09b, PI = 'all')
# Fit an alternative model assuming perceptual separability
# and perceptual independence
> PS.PI.mod.b = fit.grt(silbert09b, PS_x = T, PS_y = T, PI = 'all')
# conduct likelihood ratio tests for nested model pairs
> anova(PS.mod.b, full.mod.b)
           G2        df DG2     df p-val
PS.mod.b   11305.49  4
full.mod.b 11251.863 0  53.627  4  0
> anova(PI.mod.b, full.mod.b)
           G2        df DG2     df p-val
PI.mod.b   11315.576 4
full.mod.b 11251.863 0  63.713  4  0
> anova(PS.PI.mod.b, PS.mod.b)
            G2        df DG2     df p-val
PS.PI.mod.b 11375.917 8
PS.mod.b    11305.49  4  70.427  4  0
> anova(PS.PI.mod.b, PI.mod.b)
            G2        df DG2     df p-val
PS.PI.mod.b 11375.917 8
PI.mod.b    11315.576 4  60.342  4  0
# Check AIC for each model:
> c(full.mod.b$AIC, PS.mod.b$AIC, PI.mod.b$AIC, PS.PI.mod.b$AIC)
AIC     AIC     AIC     AIC
1757.9  1803.5  1813.6  1866.0
# Visualize the best fitting model (full.mod.b)
> plot(full.mod.b, marginals=TRUE)
```

Fig. 6. Example analysis session in R, using Gaussian GRT fitting procedures and the `silbert09b` data provided in the `mdsdt` package. All of the nested models are statistically significantly worse fitting than their more general counterparts, and the full, unconstrained model has the lowest AIC value (as well as the lowest AIC_c and BIC values), all of which indicates that this model provides the best overall account of the data. Hence, we conclude that perceptual separability and perceptual independence failed for this participant in this task. The last line produces the fitted model plot shown in Fig. 8.

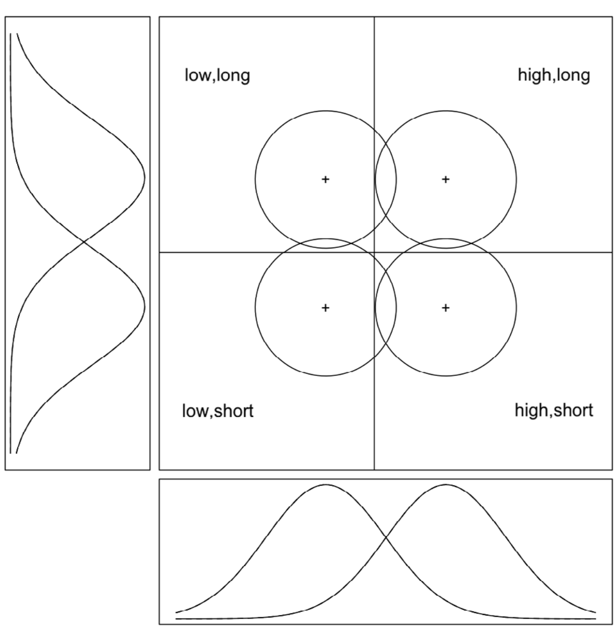

Fig. 7. Figure generated from the `plot()` function of the `mdsdt` R package, depicting the best-fitting Gaussian model for the `silbert09a` data set discussed throughout this tutorial. For this data set, model comparison results indicate that the best fitting model is the constrained model assuming perceptual separability on both dimensions and perceptual independence for all stimuli.

the first two dimensions, and the nine stimuli defining the third dimension. That is, each of these data sets consists of a $3 \times 3 \times 9$ array, where the ith 3×3 table consists of the responses to the ith stimulus. Note that the stimulus and response component on the first and second dimensions of the data will be mapped to the y and x axes, respectively, in the plotted model.

Fig. 9 illustrates a model fit to the data from Observer 1 from Thomas et al. (2015) with perceptual separability enforced on the y dimension (nose length), but not on the x dimension (eye width), and with (possible) failure of perceptual independence in all nine perceptual distributions. The function call for fitting the model in this case was `fit = fit.grt (thomas15a,PS_y=T)`, and the function call for plotting the fitted model was `plot(fit,level=0.3)`. The `level` argument may be included to shift the level at which the equal likelihood contours are plotted. Note that the level argument for the $M \times M$ model (and the concurrent ratings model; see below) specifies the *confidence* level (i.e., the volume of the perceptual distribution within the ellipse) rather than the height of the ellipse above the plane, as is the case in the 2×2 model.

Users can, of course, also analyze their own $M \times M$ factorial identification data. In order to do so, the data should be formatted as the `thomas15a` and `thomas15b` data are. That is, the data should be in a `xtabs` table, with the first two dimensions consisting of the stimulus and response components, and the third dimension consisting of the individual stimulus index. The first $M \times M$ table should correspond to the perceptual distribution at the bottom-left of the plot (i.e., A_1B_1), the second to the distribution above the first (i.e., A_1B_2), etc.,..., moving to the next column to the

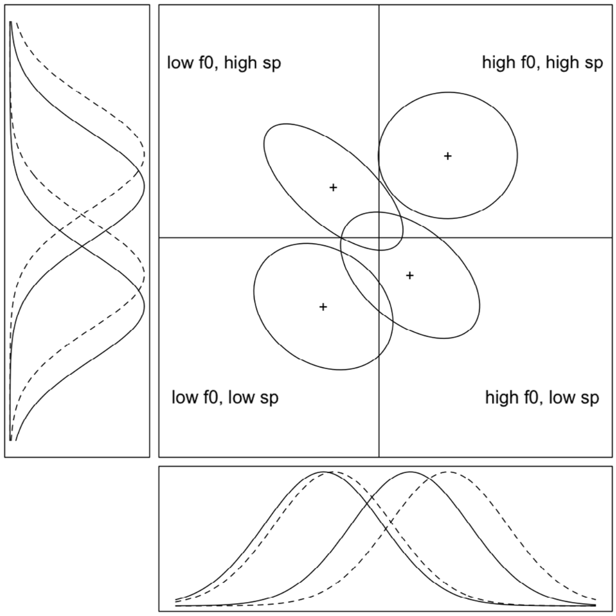

Fig. 8. Figure generated from the `plot()` function of the `mdsdt` R package, depicting the best-fitting Gaussian model for the `silbert09b` data set discussed throughout this tutorial. For this data set, model comparison results indicate that the best fitting model is the full, unconstrained model. Note that the failures of separability and independence in the Gaussian fit also nicely illustrate the results of the marginal response invariance and report independence results discussed above: the downward shift of the μ_x means at level A_1 relative to A_2 and the leftward shift of the μ_y means at level B_1 relative to B_2 illustrate the failures of marginal response invariance reported in Table 3, and the large negative correlations in stimuli A_1B_2 and A_2B_1 illustrate the failures of report independence given in Table 6.

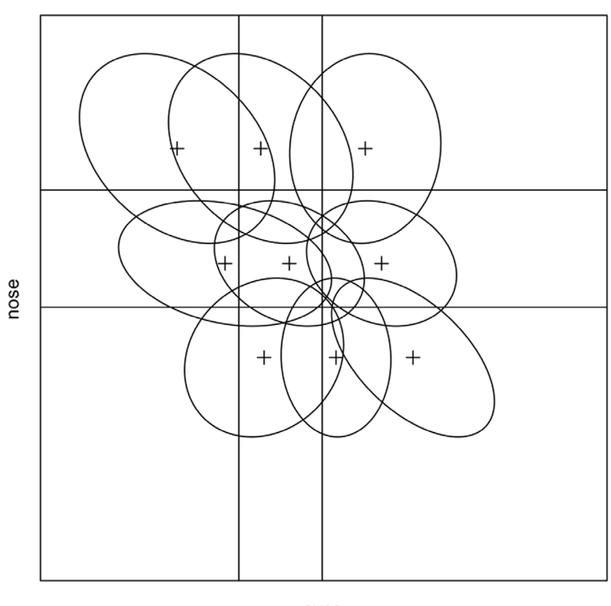

Fig. 9. Fitted 3 × 3 model for Observer 1 (Thomas et al., 2015) with perceptual separability enforced on the *y* dimension (nose length), no perceptual separability on the *x* dimension (eye width), and no perceptual independence in any perceptual distribution.

right as the highest B_j component is reached for each level of the A_i component.

It is also worth noting that the Newton–Raphson method for fitting Gaussian GRT models to data was originally developed for 'concurrent ratings' tasks and data, in which observers provide separate ratings for each dimension (Wickens, 1992). The concurrent ratings task may be carried out with the same stimulus specification and response options as the factorial identification task, but it also allows for a larger number of response levels than stimulus levels on each dimension (Ashby, 1988; Wickens & Olzak, 1989, 1992). For example, a 2 × 2 factorial stimulus set could be used for a concurrent ratings task in which $k > 2$ confidence levels are given on each dimension. The model for concurrent ratings tasks consists of a perceptual distribution for each stimulus and however many response criteria on each dimension are required to model the k response levels. The additional structure in the data and model allow for marginal variances *and* means to be estimated in all but one perceptual distribution.[11]

The `fit.grt()` function in the `mdsdt` package will estimate parameters for concurrent ratings data, and there is a concurrent ratings data set included in the package. The interested reader can import this data set by typing `data(‘‘wo89xt’’)`; this creates a `xtabs` table containing the data from Table 1 in Wickens and Olzak (1989). These data represent concurrent ratings for the absence or presence of high (*x* dimension) and low (*y* dimension) frequency components in compound sinusoid spatial gratings.[12] The concurrent ratings data should be organized such that the first dimension (rows) correspond to the response levels on the *y* axis, the second dimension (columns) corresponds to the response levels on the *x* axis, and the third dimension (layers) corresponds to the stimuli (in the order $A_1B_1, A_1B_2, A_2B_1, A_2B_2$). This data object can be given to `fit.grt()` along with any of the additional arguments specifying model constraints discussed above, and the fitted model can be then given to the `plot()` function to produce a figure illustrating the equal likelihood contours and response criteria (but not the marginal perceptual distributions). Fig. 10 illustrates the fitted model with no perceptual separability or perceptual independence enforced.

It is beyond the scope of this tutorial, but it is also worth noting that GRT has been adapted and applied to a variety of other experimental paradigms as well. As mentioned in the introduction, there is a large literature on the mathematical theory and use of GRT-based models to analyze categorization (e.g., Ashby, 1992; Ashby & Gott, 1988; Ashby & Maddox, 1993; Maddox & Ashby, 1993). GRT has also been adapted to model response times (e.g., Ashby, 1989, 2000; Townsend et al., 2012), same–different judgments (Thomas, 1996), and multidimensional structure in memory (e.g., DeCarlo, 2003). In recent work, multilevel Gaussian GRT models have been used to analyze perceptual interactions at group- and individual-listener levels (Silbert, 2012, 2014) and to fit models to multiple subjects' data by imposing strong restrictions on shared perceptual representations while relaxing the assumption of decisional separability (Soto & Ashby, 2015; Soto et al., 2104).[13]

6. Conclusion

General recognition theory provides powerful tools for analyzing interactions between dimensions in perception, memory, and

[11] It is still necessary to fix the means and marginal variances in one distribution to set the location and scale of the model as a whole.

[12] The data reported in Table 1 of Wickens and Olzak (1989) has 6 response levels, whereas the data included in the `mdsdt` package has 4. Level 1 in this table corresponds to level 1 in the wo89xt data; levels 2 and 3 in the table are collapsed to level 2 in the wo89xt data; levels 3 and 4 in the table are collapsed to level 3 in the wo89xt data; and level 6 in the table corresponds to level 4 in the wo89xt data.

[13] As noted above, this model is mathematically equivalent to a model in which decisional separability holds, and so does not represent a solution to the problem of non-identifiability of failures of decisional separability (Silbert & Thomas, submitted for publication). If independent evidence supporting the strong restrictions on the shared perceptual representations can be found, this approach may provide a solution to this problem.

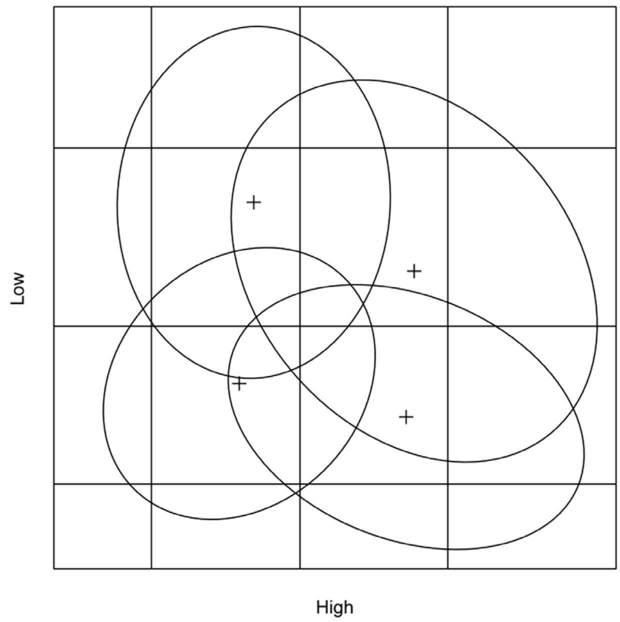

Fig. 10. Fitted 3 × 3 model for the data in Table 1 of Wickens and Olzak (1989), with perceptual separability allowed to fail on each dimension and no perceptual independence in any perceptual distribution. The y dimension corresponds to the presence/absence of a low frequency spatial grating, the x dimension to the presence/absence of a high frequency spatial grating.

other cognitive processes. However, the use of GRT has typically required a large amount of technical knowledge and a substantial time commitment. This tutorial is aimed at providing the interested reader with the technical knowledge required to understand GRT as well as the basic software tools to put GRT to use in practice.

As discussed in detail above, GRT assumes noisy encoding and deterministic response selection. In this framework, interactions between dimensions may occur (or not) within perceptual distributions (*perceptual independence*), between perceptual distributions (*perceptual separability*), or with respect to decision making (*decisional separability*). However, the presence or absence of such interactions cannot be observed directly. These latent constructs must be probed indirectly through various properties of factorial identification or concurrent ratings data.

The R package `mdsdt` provides functions for carrying out non-parametric (`mriTest()`, `riTest()`) and parametric (`fit.grt()`) analyses of such data. While requiring only very weak assumptions about the underlying processes, tests of marginal response invariance license inferences with respect to perceptual separability, while tests of report independence license inferences with respect to perceptual independence.[14] On the other hand, if the analyst is willing to make stronger assumptions about the functional form of the latent perceptual distributions, full model fitting licenses inferences about perceptual separability and independence through model comparisons (either via likelihood ratio tests or comparison of AIC, AIC$_c$, or BIC fit statistics) and allows the analyst to clearly visualize fitted models.

Appendix. Details of the parameter estimation algorithm

Although the model fitting functions are readily used without detailed knowledge of the parameter estimation algorithm, we present these details here for the sake of thoroughness. Formally, parameters in the Gaussian model are estimated using the Newton–Raphson maximum likelihood estimation algorithm described by Wickens (1992). The target is to find the set of parameters for which the response frequencies x_{ik} (i.e., the confusion matrix) is most probable. The likelihood function for the overall set of counts x_{ik} in the data has a product-multinomial form:

$$\mathcal{L}(x_{ik}|\pi_{ik}) = \prod_k \frac{n_k!}{\prod_i x_{ik}!} \prod_i \pi_{ik}^{x_{ik}}$$

where n_k is the number of times stimulus k was presented and π_{ik} is the predicted probability of responding i given stimulus k. In the feature-complete factorial designed discussed above, the outer product is taken over the four stimuli $k \in \{1, 2, 3, 4\}$ and the inner product is over the four responses $i \in \{1, 2, 3, 4\}$. In other experimental paradigms (e.g., a concurrent-ratings task; Ashby, 1988; Wickens, 1992; Wickens & Olzak, 1992) and feature-complete factorial designs with more than two dimensions and/or more than two levels on each dimension, the range of the products in the likelihood function should be adjusted appropriately.

The role of the Gaussian model is to generate the probabilities π_{ik}. Our target is to find the set of parameter vectors $\theta_k = (\mu_x^{(k)}, \mu_y^{(k)}, \rho_k)$ for which the resulting set of π_{ik} jointly maximize $\mathcal{L}$.[15] We calculate π_{ik} from the Gaussian model by integrating the Gaussian density function specified by θ_k over response region i. For example,

$$\pi_{11} = \int_{-\infty}^{c_x} \int_{-\infty}^{c_y} f(x, y|\mu_x^{(1)}, \mu_y^{(1)}, \rho_1)\, dy\, dx$$

with response region 1 corresponding to quadrant $(-\infty, c_x) \times (-\infty, c_y)$.

Typically, maximum-likelihood approaches proceed by taking the derivative of the (log) likelihood function with respect to the parameter of interest and finding critical points. However, because the Gaussian CDF cannot be represented in closed form as a function of μ_x, μ_y, and ρ, we must use a gradient ascent procedure to find its maxima. Let $\ell(\theta)$ denote the log of the likelihood $\mathcal{L}$, with probabilities π_{ik} generated from a Gaussian distribution indexed by a 1×12 vector of parameter values θ. Now, we use the standard Newton–Raphson method to find zeros of the derivative of ℓ. We can approximate this first derivative function by the first two terms in the power series around our current estimate $\theta^{(t)}$:

$$\ell'(\theta) = \ell'(\theta^{(t)}) + \ell''(\theta^{(t)})(\theta - \theta^{(t)}).$$

The Newton–Raphson procedure sets this equal to 0 and solves for θ, which is provably closer to the true maximum value than $\theta^{(t)}$. This gives us the formula

$$\theta^{(t+1)} = \theta^{(t)} - \ell''^{-1}(\theta^{(t)})\ell'(\theta^{(t)}). \quad \text{(A.1)}$$

In practice, we replace $\ell''(\theta)$ with its expectation, which is the Fisher information matrix, $\mathcal{I}(\theta)$. To evaluate this formula, we must compute expressions for the first- and second-order partial derivatives.

Let γ_{ik} denote the numerically evaluated Gaussian probability of stimulus k integrated over quadrant i. Then it can be shown (Wickens, 1992) that the partial derivatives composing $\ell'(\theta)$ are given by

$$\frac{\partial \ell}{\partial \theta_k} = \sum_{i=1}^{4} \frac{x_{ik}}{\gamma_{ik}} v_{ik}^{\theta} \quad \text{(A.2)}$$

[14] As mentioned above, failures of decisional separability are not, in general, identifiable in factorial identification or concurrent ratings data (Silbert & Thomas, 2013).

[15] Wickens (1992) algorithm was originally designed for estimating distribution and decision bound parameters in a concurrent ratings model. For the simple 2 × 2 factorial model, treating the decision bounds as the coordinate axes simplifies and stabilizes the algorithm. In this case, only perceptual distribution parameters are estimated.

and that the components of $\mathcal{I}(\theta)$ are given by

$$\mathbb{E}\left[\frac{\partial^2 \ell}{\partial\theta_k \partial\phi_k}\right] = -X_k \left[\sum_{i=1}^{4} \frac{v_{ik}^{(\theta)} v_{ik}^{(\phi)}}{\gamma_{ik}}\right] \tag{A.3}$$

where $v_{ik}^{(\theta)}$ denotes the partial derivative of γ_{ik} with respect to parameter θ. Further derivations show that for $\theta = \mu_x$, we have

$$v_{ik}^{\theta} = \begin{cases} -\phi(\mu_x^{(k)})\Phi\left(\dfrac{\rho_k \mu_x^{(k)} - \mu_y^{(k)}}{\sqrt{1-\rho_k^2}}\right) & i = 1 \\ -\phi(\mu_x^{(k)}) & i = 2 \\ 0 & i = 3, 4. \end{cases}$$

For $\theta = \mu_y$, we have

$$v_{ik}^{\theta} = \begin{cases} -\phi(\mu_y^{(k)})\Phi\left(\dfrac{\rho_k \mu_y^{(k)} - \mu_x^{(k)}}{\sqrt{1-\rho_k^2}}\right) & i = 1 \\ -\phi(\mu_y^{(k)}) & i = 3 \\ 0 & i = 2, 4 \end{cases}$$

and for $\theta = \rho$, we have

$$v_{ik}^{\theta} = \begin{cases} \phi(-\mu_x^{(k)}, -\mu_y^{(k)}, \rho_k) & i = 1, 4 \\ -\phi(-\mu_x^{(k)}, -\mu_y^{(k)}, \rho_k) & i = 2, 3 \end{cases}$$

where $\phi(c)$ denotes the standard normal density function evaluated at c and $\Phi(c)$ denotes the standard normal distribution function evaluated at c. We can insert these functions into Eqs. (15) and (16) to get a fully specified update rule. We iterate this rule (given in Eq. (14)) until the distance between successive estimates is below a given tolerance. Minor adjustments to the algorithm can be used to avoid local minima, including a "momentum" parameter w weighting the adjustment term, which gets cut in half at each iteration:

$$\theta^{(t+1)} = \theta^{(t)} - w^{(t)}\mathcal{I}(\theta^{(t)})\ell'(\theta^{(t)}).$$

Assumptions of perceptual separability and perceptual independence translate into constraints on the parameters of Gaussians. For example, perceptual separability holds on the x dimension if $\mu_x^{(1)} = \mu_x^{(2)}$ and $\mu_x^{(3)} = \mu_x^{(4)}$, and perceptual independence holds within stimulus k if $\rho_k = 0$. This involves more complex bookkeeping, but leaves the algorithm essentially unchanged: we simply sum across the different k which share a common parameter in Eqs. (15) and (16).

References

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics*, *44*(3), 195–204.

Ashby, F. G. (1989). Stochastic general recognition theory. In *Human information processing: Measures, mechanisms, and models* (pp. 435–457). Elsevier Science Publishers B. V..

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc..

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*(2), 310–329.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*(2), 150–172.

Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In *Foundations of perceptual theory* (pp. 369–399).

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 50–71.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372–400.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154–179.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*(3), 292–303.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Robert E. Krieger.

Kadlec, H., & Hicks, C. L. (1998). Invariance of perceptual spaces and perceptual separability of stimulus dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 80–104.

Kadlec, H., & Townsend, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology*, *36*(3), 325–374.

Kadlec, H., & Townsend, J. T. (1992b). Signal detection analyses of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–227). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc..

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Lawrence Erlbaum.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.

Olzak, L. A., & Wickens, T. D. (1997). Discrimination of complex patterns: orientation information is integrated across spatial scale; spatial-frequency and contrast information are not. *Perception*, *26*, 1101–1120.

Plack, C. J., & Carlyon, R. P. (1995). Loudness perception and intensity coding. In *Handbook of perception and cognition*, *Hearing* (2nd ed.) (pp. 123–160). San Diego: Academic Press.

Silbert, N. H. (2010). *Integration of phonological information in obstruent consonant identification*. (Ph.D. thesis), Indiana University.

Silbert, N. H. (2012). Syllable structure and integration of voicing and manner of articulation information in labial consonant identification. *The Journal of the Acoustical Society of America*, *131*(5), 4076–4086.

Silbert, N. H. (2014). Perception of voicing and place of articulation in labial and alveolar English stop consonants. *Laboratory Phonology*, *5*(2), 289–335.

Silbert, N. H., de Jong, K. J., Thomas, R. D., & Townsend, J. T. (2009). Diagonal d' does not (always) diagnose failure of separability: An addendum to Kingston, Diehl, Kirk, and Castleman (2008). *Journal of Phonetics*, *37*(3), 339–343.

Silbert, N. H., & Thomas, R. D. (2013). Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Bulletin & Review*, *20*, 1–20.

Silbert, N. H., & Thomas, R. D. (2014). Optimal response selection and decisional separability in Gaussian general recognition theory. *Journal of Mathematical Psychology*, *60*, 72–81.

Silbert, N. H., & Thomas, R. D. (2016). Model identifiability in GRT with individual differences. *Journal of Mathematical Psychology*, (submitted for publication).

Silbert, N. H., Townsend, J. T., & Lentz, J. J. (2009). Independence and separability in the perception of complex nonspeech sounds. *Attention, Perception, & Psychophysics*, *71*(8), 1900–1915.

Soto, F. A., & Ashby, F. G. (2015). Categorization training increases the perceptual separability of novel dimensions. *Cognition*, *139*, 105–129.

Soto, F. A., Vucovich, L., Musgrave, R., & Ashby, F. G. (2014). General recognition theory with individual differences: a new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin & Review*, *22*(1), 88–111.

Thomas, R. D. (1995). Gaussian general recognition theory and perceptual independence. *Psychological Review*, *102*(1), 192.

Thomas, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology*, *40*(4), 318–341.

Thomas, R. D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general d. *Psychonomic Bulletin & Review*, *6*(2), 224–238.

Thomas, R. D. (2001a). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 193–228).

Thomas, R. D. (2001b). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics*, *63*(4), 625–650.

Thomas, R., Altieri, N., Silbert, N., Wenger, M., & Wessels, P. (2015). Multidimensional signal detection decision models of the uncertainty task: Application to face perception. *Journal of Mathematical Psychology*, *66*, 16–33.

Thomas, R. D., & Silbert, N. H. (2014). Technical clarification to Silbert and Thomas (2013): Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Bulletin & Review*, *21*(2), 574–575.

Townsend, J. T., Houpt, J. W., & Silbert, N. H. (2012). General recognition theory extended to include response times: Predictions for a class of parallel systems. *Journal of Mathematical Psychology*, *56*(6), 476–494.

Wickens, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology*, *36*(2), 213–234.

Wickens, T. D., & Olzak, L. A. (1989). The statistical analysis of concurrent detection ratings. *Perception & Psychophysics*, *45*(6), 514–528.

Wickens, T. D., & Olzak, L. A. (1992). Three views of association in concurrent detection ratings. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (p. 523). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.