

**INTEGRATION OF PHONOLOGICAL INFORMATION
IN OBSTRUENT CONSONANT IDENTIFICATION**

Noah H. Silbert

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Linguistics
& the Department of Cognitive Science
Indiana University
May, 2010

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Kenneth J. de Jong, Ph.D.

James T. Townsend, Ph.D.

Jennifer J. Lentz, Ph.D.

David B. Pisoni, Ph.D.

September 21, 2009

Dedicated to Rejoyce Enright, whose patience knows only the most expansive of bounds.

Acknowledgements

This thesis would not exist without the hard work, time, and patience of a number of people other than me. Over the course of many years, Ken de Jong taught me much of what I know about phonology and phonetics, Jim Townsend taught me much of what I know about mathematical modeling and psychology, and Jenny Lentz taught me much of what I know about psychoacoustics and auditory experimentation. Other people who have played any of a number of important roles in my education include (in alphabetical order by surname) Dan Blaine, Robert Botne, J. D. Brown, Dan Dinnsen, Cathy Doughty, Rob Goldstone, Thom Hudson, Diane Kewley-Port, John Kruschke, Mike Long, Rob Nosofsky, David Pisoni, Bob Port, and Herb Roitblat. Of course, my mother and father gave me the genes and, with my brother, an upbringing that turns out to have been sufficient to the task of getting me, in order of occurrence, an MA, a PhD, and an MS (forthcoming). Last but not least, my wife and children daily suffer the travails of cohabitation with an academic.

Abstract

Speech perception requires the integration of information from multiple phonetic and phonological dimensions. Numerous studies have investigated the mapping between multiple acoustic-phonetic dimensions and single phonological dimensions (e.g., spectral and temporal properties of stop consonants in voicing contrasts). Many fewer studies have addressed relationships between phonological dimensions. Most such studies have focused on the perception of sequences of phones (e.g., [bɪd], [bɛd], [brɪ], [bɛt]), though some have focused on multiple phonological dimensions within phones (e.g., voicing and place of articulation in [p], [b], [t], and [d]). However, strong assumptions about relevant acoustic-phonetic dimensions and/or the nature of perceptual and decisional information integration limit previous findings in important ways. New methodological developments in the General Recognition Theory framework enable a number of these assumptions to be tested and provide a more complete model of distinct perceptual and decisional processes in speech sound identification. A Bayesian non-parametric analysis of data from four experiments probing identification of (two sets of) consonants in onset (syllable initial) and coda (syllable final) position indicate that integration of phonological information is partially independent in both perception and decision making for most subjects, and that patterns of independence and interaction vary with the set of phonological dimensions under consideration and with syllable position.

Contents

Chapter 1. Multiple dimensions in speech perception	1
1.1. Introduction and motivation	1
1.2. Category structure on a single phonological dimension	2
1.3. Category structure on multiple phonological dimensions	3
1.4. Limitations of previous work	5
1.5. Roadmap	6
Chapter 2. General Recognition Theory	8
2.1. The structure of General Recognition Theory	8
2.2. Three forms of independence	9
2.3. Testing independence	11
Chapter 3. Statistical inference, marginal response invariance, and report independence	17
3.1. Tests of Marginal Response Invariance and Report Independence	17
3.2. The Bayesian model	19
3.3. Relationships between known GRT models, Δ , and π	23
3.4. Summary	30
Chapter 4. Place of articulation and voicing in onset position	31
4.1. Methods	31
4.2. Results	33
4.3. Discussion	44
Chapter 5. Place of articulation and voicing in coda position	46
5.1. Methods	46
5.2. Results	47
5.3. Discussion	57

Chapter 6. Manner of articulation and voicing in onset position	60
6.1. Methods	61
6.2. Results	62
6.3. Discussion	70
Chapter 7. Manner of articulation and voicing in coda position	73
7.1. Methods	74
7.2. Results	74
7.3. Discussion	82
Chapter 8. Conclusion	85
8.1. General Recognition Theory and a multinomial-Dirichlet statistical model	85
8.2. Recapitulation of experimental results	87
8.3. Possible Causes and Implications	91
8.4. Limitations	93
8.5. Future work	95
8.6. Closing remarks	99
References	100
Appendix A. Statistical power in the null-hypothesis tests and Bayesian model	105
A.1. Perceptual Independence, Decisional Separability, and Report Independence	105
A.2. Perceptual Separability, Decisional Separability, and Marginal Response Invariance	108
Appendix B. BRUGS Code	111
Appendix C. Acoustic measurements of stimuli	113

Multiple dimensions in speech perception

1.1. Introduction and motivation

As a speech signal is mapped onto a meaningful linguistic representation, information from multiple acoustic-phonetic and phonological dimensions must be integrated. Consider, for example, an utterance beginning with the voiceless bilabial stop [p]. At any given point in time, the signal will be perceived as more or less [p]-like and more or less [b]-like; the percept will be more or less (phonologically) voiced. It will be perceived as more or less [t]-like and more or less [f]-like, as well; the percept will be more or less labial or alveolar, more or less a good example of a stop or fricative.

In the absence of noise, a clearly articulated [p] will be perceived as far more [p]-like than [b]-, [t]-, or [f]-like, though in many common settings, the signal will be more ambiguous. In any case, in order to accurately perceive a spoken [p], information from voicing, place of articulation, and manner of articulation dimensions (among others) must be integrated. It may be that these phonological dimensions are integrated independently; a phoneme may be a simple combination of its component parts. On the other hand, it may be that phonological dimensions interact, producing a percept that is qualitatively different than would be expected from a simple combination of its component parts.

Investigations of the integration of phonological information are complicated by the fact that the mapping between acoustic-phonetic dimensions and phonological dimensions is many-to-many. For example, both stop burst strength and voice onset time (VOT) cue voicing distinctions, but both also play a role in cuing place of articulation. For example, more intense release bursts and longer VOTs cue, in English, a voiceless stop, but voiceless alveolar stops tend to have more intense release bursts and longer VOTs than do voiceless labial stops (e.g., Oglesbee, 2008).

Previous studies probing the perception of distinctive features and their acoustic-phonetic correlates have been limited by strong assumptions about the nature of the relationships between dimensions, a failure to consider all the levels at which dimensions may interact, or both. These

limitations may be addressed directly in the General Recognition Theory (GRT) framework (Ashby & Townsend, 1986; Kadlec & Townsend, 1992a; Thomas, 2001a), a multidimensional generalization of signal detection theory (SDT; Green & Swets, 1966). GRT is well suited to this purpose because it provides rigorous distinctions between independence of dimensions in perception and decision making and between within-stimulus and across-stimuli levels of interaction.

We turn now to a discussion of previous work on the internal (perceptual) structure of pairs of contrasting phonological segments (i.e., the structure of phonological categories with regard to a single phonological dimension). This is followed by a discussion of research in which the relationships between multiple phonological dimensions have been directly addressed. This chapter ends with a brief ‘roadmap’ describing the organization of the rest of the thesis.

1.2. Category structure on a single phonological dimension

As noted briefly above, a phonological contrast may have multiple acoustic correlates, and different phonological contrasts may ‘share’ acoustic distinctions. For example, voicing in syllable onset position can be cued by VOT (e.g., Lisker & Abramson, 1964; Stevens, 1998), though a number of other acoustic-phonetic properties also distinguish voiced from voiceless stop consonants, such as consonant release-burst amplitude, aspiration noise amplitude, and fundamental frequency (F0) at voice onset (e.g., Oglesbee, 2008). Similarly, voicing distinctions in onset fricatives correspond to differences in frication noise amplitude and duration, as well presence or absence of voicing during noise production (e.g., Silbert & Jong, 2008; Stevens, 1998).

Place of articulation distinctions also map onto multiple acoustic-phonetic properties. In stop consonants, place of articulation can be cued by differences in the spectral structure and amplitude of the release-burst, whereas in fricatives, place can be cued by differences in the spectral structure and duration of frication noise. The location and movement of vowel formants near the consonant-vowel boundary also vary as a function of place in both stops and fricatives (as well as other consonants; Stevens, 1998).

Not surprisingly, the internal perceptual structure of phonological categories seems to depend, at least in part, on the nature of the relevant acoustic-phonetic properties. For example, identification accuracy and category goodness judgments for stop consonants vary as a function of VOT, and this variation depends on speaking rate and place of articulation. Relative to faster speaking rates,

slower speaking rates broaden and shift (upward) the range of VOTs judged as good exemplars of voiced and voiceless categories (J. L. Miller & Volaitis, 1989; Volaitis & Miller, 1992).

However, while VOT is an important cue to voicing distinctions, it is not the only acoustic-phonetic dimension relevant to this contrast. The frequency of F0 and F1 at voicing onset and the amplitude of the release burst also play a role in distinguishing voiced from voiceless obstruents in onset position (Kingston & Diehl, 1994). Recent work probing a large number of acoustic-phonetic dimensions supports these claims and indicates further that the internal structure of phonological categories also varies as a function of a listener's native language (e.g., Oglesbee, 2008).

In a few cases, research on the internal structure of phonological categories has focused explicitly on interactions between acoustic-phonetic dimensions. It has been shown convincingly that F1 frequency and spectral correlates of nasalization interact perceptually (Kingston & Macmillan, 1995; Macmillan, Kingston, Thorburn, Dickey, & Bartels, 1999), as do voice quality and F1 frequency (Kingston, Macmillan, Dickey, Thorburn, & Bartels, 1997). More recently, perceptual interactions between a number of spectral and temporal cues to voicing in intervocalic consonants have been reported (Kingston, Diehl, Kirk, & Castleman, 2008).

1.3. Category structure on multiple phonological dimensions

Relationships between phonological dimensions have been probed directly in a smaller number of studies. It has been found, for example, that slightly longer VOTs are judged as better for velar place of articulation than for labial place of articulation (Volaitis & Miller, 1992). Evidence of interactions between place and voicing was also presented by Sawusch and Pisoni (1974). Similar shifts toward longer VOTs for alveolar stops relative to labial stops in categorization have been shown to correspond closely to statistical information from category structure in production (i.e., the distribution of VOTs; Nearey & Hogan, 1986). Interactions between place of articulation and voicing, place and manner of articulation, and a number of segmental and suprasegmental variables have been found in speeded classification tasks (Eimas, Tartter, Miller, & Keuthen, 1978; Eimas, Tartter, & Miller, 1981; J. L. Miller, 1978). On the other hand, indirect evidence of independence between phonological dimensions has been found in phonetic categorization (J. L. Miller & Eimas, 1977), suggesting that patterns of interaction and independence may depend on the experimental task employed.

However, although the results of Volaitis and Miller (1992) and Eimas et al. (1978) provide reasonably clear evidence of interactions between (phonological) dimensions, neither maintains a rigorous distinction between perceptual and decisional processes. To the extent that perception and decision making are distinct cognitive processes, this makes interpretation of evidence of interaction between phonological dimensions difficult.

A small body of literature exists in which both perceptual and decisional interactions between phonological dimensions are explicitly addressed. The Normal *A Posteriori* Probability model (Nearey & Hogan, 1986; Nearey, 1990) explicitly models both the structure of perceptual categories and the decision making process. Categorization of sequences of phones has also been modeled using logistic regression models with decision bias parameters free to vary without regard to category structure per se (Nearey, 1992, 1997, 2001, 2003) and, with closely related mathematical underpinnings, as a hierarchical process in which decisions about one dimension (may) influence decisions about another dimension (Smits, 2001a, 2001b).

Fits of the HICAT (Hierarchical CATegorization) model to three sets of categorization data (two sets from fricative-vowel stimuli, and one set from vowel-stop stimuli) support the idea that decision making on one phonological dimension can depend on the perception of another dimension (Smits, 2001a, 2001b). For example, HICAT-based analyses indicate that the boundary separating the alveolar fricatives [s] and [ʃ] varies as a function of the perception of the following vowel ([i] or [u] in one data set, [i] or [y] in another) and that the boundary separating the stops [d] and [t] varies as a function of the perception of a preceding vowel ([e^h] or [ɛ]). Perceptual dimensions in HICAT are assumed to correspond directly to acoustic-phonetic dimensions.

Perception and decision making are also addressed as distinct mechanisms in the fuzzy logical model of perception (FLMP), which has been used to model integration of information from various acoustic-phonetic dimensions (e.g., VOT, formant values at voicing onset, aspiration noise intensity) in the categorization of speech sounds varying on multiple phonological dimensions (voicing and place of articulation in nonsense syllable initial consonants [pæ], [bæ], [tæ], [dæ], as in Massaro & Oden, 1980; Oden & Massaro, 1978). Equivocal evidence of decisional interactions and perceptual independence was found in at least one FLMP-based analysis of categorization data (Oden & Massaro, 1978).

1.4. Limitations of previous work

The work described above makes it clear that there is rich internal structure to phonological categories. However, the waters are muddied by a number of theoretical and methodological limitations to the evidence of interactions between phonological dimensions in much previous work. We consider each in turn.

1.4.1. Theoretical limitations. There are two major theoretical limitations to previous work. First, though interactions between dimensions may occur at multiple, distinct levels (e.g., perception and decision making), these levels are often conflated. Second, strong assumptions about the presence or absence of interactions at one or more levels are often made *a priori*.

1.4.1.1. *Conflation of levels of interaction.* It is clear that explicit goodness judgments are the product of both perceptual and decisional factors; the results reported by Volaitis and Miller (1992) could be due to either or both in conjunction. Similarly, Maddox (1992) has shown that interference in speeded classification (as used by, e.g., Eimas et al., 1981) could be the result of either perceptual or decisional processes. Insofar as the structure and location of categories is only implicitly considered in, e.g., logistic regression models of categorization (e.g., Nearey, 1992) or the FLMP (e.g., Oden & Massaro, 1978), within- and across-category forms of perceptual interaction are conflated.

The distinctions between perception and decision making and between within-stimulus and across-stimuli levels of interactions (or independence) are important to maintain. On one hand, decision making is logically distinct from perception, taking place post-perceptually and at least partially under the control of listeners. Dimensional interactions in decision making implicate ‘higher-level’ cognitive systems than do perceptual interactions. On the other hand, across-stimuli perceptual interactions indicate imperfect generalization of one (value of one) dimension across levels of another (e.g., voicing across levels of place of articulation), whereas within-stimulus perceptual interactions provide information about the ‘shape’ of categories defined by particular levels of multiple dimensions (e.g., voiced labials).

1.4.1.2. *Assumptions of independence.* A complete picture of the interactions between phonological dimensions should rely on as few *a priori* assumptions about the presence or absence of interactions as possible. While within-stimulus, across-stimuli, and decisional notions of independence are logically distinct (Ashby & Townsend, 1986), they may well covary in reality (e.g., Ashby

& Lee, 1991). The assumption that one form of independence (e.g., within-stimulus independence) holds has incompletely understood implications for investigations into other forms of independence (e.g., between-stimuli perceptual and decisional independence). Assumptions of within-stimulus independence (i.e., absence of interactions) have been employed in a number of previous studies.

To the best of our knowledge, all previous applications of multidimensional detection theory to speech perception have assumed within-stimulus independence (e.g., Kingston & Macmillan, 1995; Kingston et al., 1997, 2008; Macmillan et al., 1999). Similarly, in the FLMP, independent integration of information from multiple acoustic dimensions is assumed (Massaro & Oden, 1980; Oden & Massaro, 1978). On the other hand, in a number of applications of GRT to visual perception, the decisional notion of independence is often assumed to hold (e.g., Olzak & Wickens, 1997; Thomas, 2001a, 2001b).

1.4.2. Methodological limitations. Evidence of interaction between phonological dimensions may also be limited in scope simply because it arises as an unintended consequence of work on other issues. For example, although the results of both Nearey and Hogan (1986) and Volaitis and Miller (1992) indicate that perception of (and/or decision making with regard to) voicing varies as a function of place of articulation, neither provides information regarding perception of place of articulation as a function of voicing. This can be readily addressed by requiring complete identification of each level of each dimension of interest (Ashby & Townsend, 1986).

As discussed above, much evidence of phonological dimension interaction comes from work focused on relationships between acoustic-phonetic dimensions in categorization (e.g., Nearey, 1992; Smits, 2001b). Most such studies employ stimuli built on a number of acoustic-phonetic continua defined on predetermined dimensions (e.g., VOT, formant frequency value at voice onset, etc...). Although this method has proven value, if the goal is to study the relationships between phonological dimensions (as it is here), we should avoid strong assumptions about the relevant acoustic-phonetic structure of the stimuli insofar as we are able to do so. This may be accomplished by employing naturally produced stimuli, for example.

1.5. Roadmap

The present work employs non-parametric GRT in an investigation of interactions between phonological dimensions in obstruent consonants. GRT defines three levels at which interactions (or

independence) between dimensions may obtain, thereby avoiding the problem of conflation of levels of interaction. The non-parametric measures of interaction introduced by Ashby and Townsend (1986) enable analyses of interactions between dimensions without making strong assumptions about particular forms of independence holding *a priori*; these non-parametric measures are employed here in a Bayesian statistical framework. Finally, naturally produced tokens of nonsense syllables containing the consonants of interest were employed as stimuli so as to avoid making strong assumptions about the relevant acoustic-phonetic dimensions.

The rest of the thesis is organized as follows. Chapter 2 describes GRT in detail. Chapter 3 provides a discussion of statistical inference, measures of interaction between dimensions, and presents a hierarchical Bayesian model used here to link data, measurements, and theory. Chapter 4 presents an analysis of interactions between place of articulation and voicing in onset obstruent consonants. Chapter 5 presents an analysis of interactions between place and voicing in coda consonants. Chapter 6 presents an analysis of interactions between manner of articulation and voicing in onset obstruent consonants. Chapter 7 presents an analysis of interactions between manner and voicing in coda consonants. Chapter 8 concludes with a discussion of the implications and limitations of this work, as well as some possible future directions.

General Recognition Theory

This introduction to the structure of GRT provides definitions of three logically distinct notions of independence, a description of the experimental protocols typically employed with GRT, and a brief discussion of various methods of testing the various forms of independence.

2.1. The structure of General Recognition Theory

As noted above, GRT is a multidimensional generalization of SDT. As such, it is a two-stage model of perception and decision-making. In GRT, two major assumptions are made. First, it is assumed that the presentation of a stimulus produces a random perceptual effect due to internal (e.g., neural) noise, external noise added to the stimulus, or both. Over the course of many trials, this results in distributions of perceptual effects.¹ Second, it is assumed that perceptual space is exhaustively partitioned by decision bounds into mutually exclusive response regions. The response regions determine the response associated with (sets of) perceptual effects.

In its most general form, GRT is non-parametric, involving no assumptions about the specific functional forms of the perceptual distributions or decision bounds. This allows for rigorous, general definitions of three logically distinct notions of independence (and interaction) between dimensions. Two of these concern perception, while the other concerns decision making. One of the perceptual notions concerns within-stimulus independence, while the other concerns across-stimuli independence.

Strictly for the purposes of illustration, two additional assumptions will be made. First, we make the assumption that the perceptual distributions are bivariate Gaussian. Second, we assume that decision bounds are linear. The non-parametric analyses to be employed require neither of these assumptions. They are non-parametric precisely because no parameters are required to define

¹ The distributions in SDT and GRT need not, in general, be considered perceptual. For example, SDT has been profitably applied to memory research in which no plausible role for perceptual distributions exists, as in DeCarlo (2002). Throughout the present work, however, distributions will be assumed to be perceptual in nature.

particular functional forms for the perceptual distributions or decision bounds. Nonetheless, these additional assumptions allow a clear and simple illustration of the model.

Figure 2.1 shows an illustrative two dimensional Gaussian GRT model for the stops [p], [b], [t], and [d], i.e., consonants consisting of the factorial combination of (labial and alveolar) place of articulation and (voiced and unvoiced) voicing. It is convenient in visualization of (Gaussian) GRT models to make use of equal likelihood contours, or sets of points the same height above the (x, y) plane; here, (x, y) correspond to $(place, voicing)$. Given the presentation of a particular stimulus, the points inside the appropriate ellipse are more likely to occur (i.e., are more likely perceptual effects) than are the points outside the ellipse.

2.2. Three forms of independence

The within-stimulus notion of independence in the GRT framework is called *perceptual independence*. Perceptual independence holds for a given perceptual distribution if, and only if, stochastic independence holds. Thus, perceptual independence holds if, and only if, for all perceptual effects corresponding to a given stimulus, the probability of the joint perceptual effect of particular levels on each dimension (e.g., voiceless and labial) is equal to the product of the marginal probabilities of the perceptual effects of each level considered separately. With the additional assumption of Gaussian distributions, perceptual independence holds if, and only if, correlation between the perceptual effects on each dimension is zero.

The two bivariate distributions on the left of the main panel of Figure 2.1 illustrate perceptual independence. The equal likelihood contour for [p] is a circle because the marginal variance for place is equal to the marginal variance for voicing in this distribution, whereas the equal likelihood contour for [b] is an ellipse because the marginal variance for voicing is larger than the marginal variance for place. In each case, the correlation between place and voicing is zero. On the other hand, perceptual independence fails for the two distributions on the right. In the [t] distribution, there is negative correlation between place and voicing (i.e., the more ‘alveolar’ a perceptual effect is, the more likely it is to also be ‘voiceless’), whereas there is a positive correlation between the perceptual effects in the [d] distribution. The sign of the correlation is defined with respect to the dimensions and levels; the effect of perceptual correlations is to change the relative proportions of distributions that fall in the different response regions.

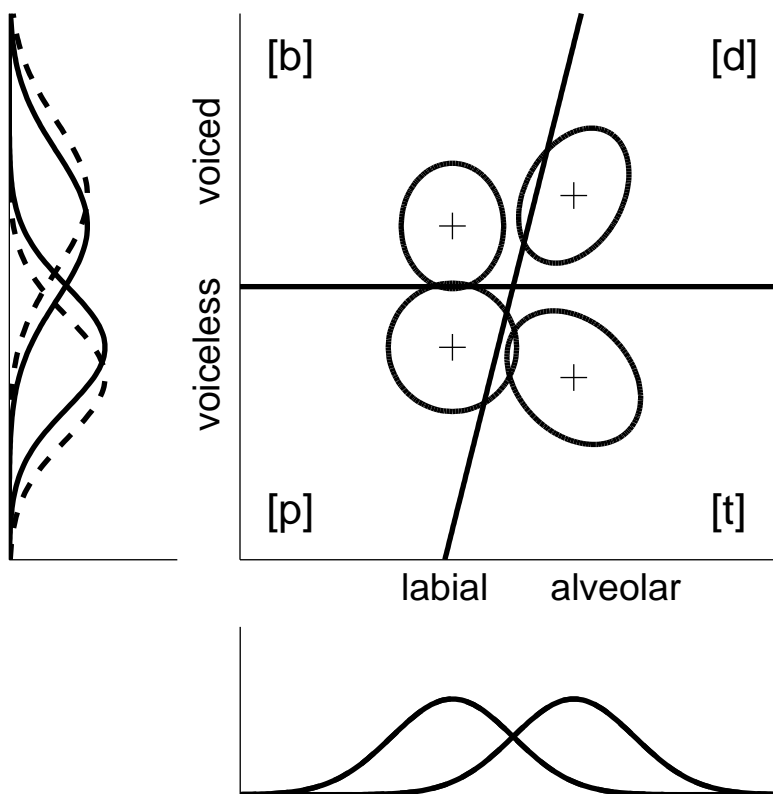


FIGURE 2.1. Illustrative two dimensional Gaussian GRT model. The ellipses represent contours of equal likelihood; the plus signs indicate the means of the distributions. The solid line marginal densities correspond to the first level on the other dimension (e.g., voiceless or labial); the dashed line marginals correspond to the second level (e.g., voiced or alveolar). See text for details.

The across-stimuli notion of independence in GRT is called *perceptual separability*. We say that, for example, place is perceptually separable from voicing if the marginal perceptual effect of place is identical across levels of voicing. In Figure 2.1, this is illustrated for place of articulation. The marginal perceptual distribution of voiced labial stops is identical to the marginal perceptual distribution of voiceless labial stops; the same pattern holds for alveolars. In this case, we would say that place of articulation is perceptually separable from voicing. On the other hand, in the model illustrated in Figure 2.1, voicing is *not* perceptually separable from place. Perceptual separability has failed due to shifts in the marginal means for alveolar stops such that voiceless alveolar stops

tend to be perceived as ‘more voiceless’ than are voiceless labial stops and voiced alveolar stops tend to be perceived as ‘more voiced’ than are voiced labial stops. Perceptual separability can also fail due to shifts in means in the same direction (e.g., the alveolar stops could both tend to be more voiced than their labial counterparts). Changes in marginal variances can, in principle, also cause failure of perceptual separability.

Finally, the decision-related notion of independence in GRT is called *decisional separability*. Decisional separability holds if, and only if, a decision bound is parallel to the (appropriate) coordinate axis. In Figure 2.1, decisional separability holds for the bound separating the voiced and voiceless response regions. In this case, decisions about the voicing category of a stimulus would not depend on either perceptual effects or decision-making about the stimulus’ place. Decisional separability fails here for the bound separating the labial and alveolar response regions such that there is a bias toward responding ‘labial’ in the voiced region of perceptual space, whereas there is a bias toward responding ‘alveolar’ in the voiceless region of perceptual space.

The simplest experimental protocol in which all three of these notions of independence can be investigated consists of identification of four stimuli defined by the factorial combination of two levels on each of two dimensions. Although GRT can be extended to an arbitrary number of levels on an arbitrary number of dimensions (see, e.g., Kadlec & Townsend, 1992b), two levels on two dimension is the most commonly used experimental and theoretical structure. For the remainder of this work, we will restrict ourselves to this ‘standard’ GRT situation. In this case, the model predicts identification-confusion probabilities as the volumes of the bivariate perceptual distributions in the appropriate response regions. For example, the predicted probability of responding [t] when presented with [p] would be the volume of the [p] distribution in the [t] response region.

2.3. Testing independence

General non-parametric² tests of independence can be carried out on observed identification confusion data (Ashby & Townsend, 1986; Thomas, 2001a, 2001b), as can (marginal) SDT-based tests of all three forms of independence (i.e., tests employing marginal d' and β parameters; Kadlec & Townsend, 1992a, 1992b). In addition, parameter estimation (i.e., model fitting) and model

² Tests that do not assume particular parametric forms of underlying perceptual distributions or decision bounds

comparison can also be used to test various forms of independence (e.g., Thomas, 2001a; Wickens, 1992).

All of these approaches have important limitations. The non-parametric tests only test conjunctions of forms of independence (i.e., perceptual independence *and* decisional separability, perceptual *and* decisional separability). The SDT-based tests can only detect certain forms of failures of independence. For example, one such test probes equality of marginal voicing d' values across levels of place. If separability fails because, say, the voicing is more salient for alveolars than for labials, this test is, in principle, able to detect such a failure of perceptual separability. If, on the other hand, perceptual separability fails because the alveolar stops are uniformly perceived as more voiceless than are the labial stops (i.e., if the alveolar distributions are shifted in the same direction along the voicing dimension), the marginal voicing d' values could be equal across levels of place despite the failure of separability (see, e.g., Kingston et al., 2008, for examples of possible failures of perceptual separability of this type).

Parameter estimation in GRT also suffers from important limitations. This is due, at least in part, to the fact that the most general Gaussian GRT model (with linear bounds) has a large number of free parameters, typically more than the data have degrees of freedom. This means that the data do not constrain the model; multiple configurations of perceptual distributions and decision bounds may account for a given data set equally well. The most common solution to this has been to assume that decisional separability holds while testing perceptual independence and perceptual separability (Olzak, 1986; Thomas, 2001a, 2001b; Wickens, 1992).

Two methods have been developed to overcome this limitation to model fitting analyses, both, at least in part, by increasing the degrees of freedom in the data relative to the model. The first employed nine stimuli defined at three levels on each of two dimensions (Ashby & Lee, 1991). The second manipulates stimulus presentation base rates (the frequency with which different stimuli are presented in an experimental block) with the intention to shift decision bounds without affecting the shape or location of perceptual distributions (Maddox, 1995; Maddox & Bohil, 1998a, 1998b; Silbert, Townsend, & Lentz, 2007a, 2007b). Although both approaches enable more general GRT models to be employed, both require additional, possibly problematic, assumptions.

The present work employs two non-parametric measures of dimensional interaction (Ashby & Townsend, 1986; Townsend, Hu, & Ashby, 1981; Townsend, Hu, & Evans, 1984) – marginal response invariance and report independence³ – in a non-parametric Bayesian statistical model. Both marginal response invariance and report independence are defined in terms of observed identification-confusion proportions. Thus, they are not tied to any specific assumptions about the functional form of the underlying perceptual distributions or decision bounds. This makes them very general measures of dimensional interaction, though, as noted above, they only test conjunctions of forms of independence.

Let capital letters correspond to stimulus attributes and lower case letters correspond to responses. Thus, for a given stimulus, A_i indicates the i^{th} level on the first dimension, and B_j indicates the j^{th} level on the second dimension. The response to the presentation of the A_i, B_j stimulus would then be denoted a_k, b_l , where $i, j, k, l = 1, 2$. A bit more more concretely, given that the stimulus at level one on the first dimension and level two on the second dimension (i.e., the A_1, B_2 stimulus) was presented, the probability of responding correctly on the first dimension and incorrectly on the second dimension would be denoted $p(a_1, b_1|A_1, B_2)$.

Marginal response invariance holds for dimension A as a whole if the equality given in equation 1 holds for $i = 1, 2$:⁴

$$(1) \quad p(a_i, b_1|A_i, B_1) + p(a_i, b_2|A_i, B_1) = p(a_i, b_1|A_i, B_2) + p(a_i, b_2|A_i, B_2)$$

Written in terms of levels on a phonological dimension, the equations defining marginal response invariance with respect to place of articulation are as follows, where $p('p'|[b])$ indicates the probability of responding 'p' when presented with [b]. Each equation is written twice, once in terms of individual consonant responses (e.g., 'p' or 'b') and once in terms of marginal, feature-level responses (e.g., 'labial').

³ also called sampling independence

⁴ Marginal response invariance may hold a single level on a given dimension.

$$\begin{aligned} p('p'|[p]) + p('b'|[p]) &= p('p'|[b]) + p('b'|[b]) \\ p('labial'|[p]) &= p('labial'|[b]) \end{aligned}$$

$$\begin{aligned} p('t'|[t]) + p('d'|[t]) &= p('t'|[d]) + p('d'|[d]) \\ p('alveolar'|[p]) &= p('alveolar'|[b]) \end{aligned}$$

Similarly, marginal response invariance holds for dimension B if the equality given in equation 2 holds for $j = 1, 2$:

$$(2) \quad p(a_1, b_j | A_1, B_j) + p(a_2, b_j | A_1, B_j) = p(a_1, b_j | A_2, B_j) + p(a_2, b_j | A_2, B_j)$$

As above, written in terms of levels on a phonological dimension, the equations defining marginal response invariance with respect to voicing are as follows:

$$\begin{aligned} p('p'|[p]) + p('t'|[p]) &= p('p'|[t]) + p('t'|[t]) \\ p('voiceless'|[p]) &= p('voiceless'|[b]) \end{aligned}$$

$$\begin{aligned} p('b'|[b]) + p('d'|[b]) &= p('b'|[d]) + p('d'|[d]) \\ p('voiced'|[p]) &= p('voiced'|[b]) \end{aligned}$$

Thus, marginal response invariance holds if the probability of correctly⁵ identifying the level on a given dimension is equal across levels of the other dimension. That is, it is a measure of

⁵ Strictly speaking, marginal response invariance need not be defined in terms of correct identification of a given level. However, if performance is above chance, there will be more data in the correct cells in the confusion matrices, which should provide more reliable estimates of response probabilities. Note, too, that in the two levels on two dimensions case assumed here, the probability of incorrectly identifying the level on a dimension is one minus the probability of correctly identifying the level on that dimension, so incorrect and correct responses should produce identical measures of marginal response invariance

generalization. As implied by the pairs of equations given above in terms of place of articulation and voicing, respectively, in the standard, two levels on each of two dimensions GRT case, marginal response invariance may hold or fail at one level regardless of whether it holds or fails at the other.

Marginal response invariance is related to GRT in Theorem 5 of Ashby and Townsend (1986), which states that the conjunction of decisional separability and perceptual separability implies marginal response invariance, but not vice versa. Thus, if marginal response invariance fails in a data set, we can conclude either that perceptual separability has failed, that decisional separability has failed, or that both have failed. If marginal response invariance holds in a data set, we cannot draw any strong inferences, though we may take it as suggesting that both perceptual separability and decisional separability hold.

Report independence holds when presented with stimulus A_i, B_j if the equality in equation 3 holds:

$$(3) \quad p(a_i, b_j | A_i, B_j) = \left[p(a_i, b_1 | A_i, B_j) + p(a_i, b_2 | A_i, B_j) \right] \times \left[p(a_1, b_j | A_i, B_j) + p(a_2, b_j | A_i, B_j) \right]$$

Written in phonological terms with respect to [p], the definition of report independence is:

$$\begin{aligned} p('p'|[p]) &= \left[p('p'|[p]) + p('b'|[p]) \right] \times \left[p('p'|[p]) + p('t'|[p]) \right] \\ p('labial, voiceless'|[p]) &= p('labial'|[p]) \times p('voiceless'|[p]) \end{aligned}$$

Thus, report independence holds given the presentation of a particular stimulus if the joint probability of issuing a correct⁶ response is equal to the product of the marginal probabilities of correctly identifying the levels of each component dimension of that stimulus. Report independence is a measure of statistical independence, and is related to GRT in Theorem 1 of Ashby and Townsend (1986), which states that the conjunction of perceptual independence and decisional separability implies report independence. Thus, if report independence fails in a data set, we can conclude either that perceptual independence has failed, that decisional separability has failed, or that both have failed. As with marginal response invariance, if report independence holds in a data set, we cannot

⁶ Again, strictly speaking, report independence need not be defined only in terms of correct responses. Nonetheless, we will do so here.

draw strong inferences, though we may take it as suggesting that both perceptual independence and decisional separability hold.

Statistical inference, marginal response invariance, and report independence

The problem of statistical inference with regard to marginal response invariance, report independence, and perceptual and decisional interactions is threefold. First, we desire a method for reliably detecting interactions; we want a measure of interaction and a rule for deciding if the measure accurately indicates the presence of interaction(s). Second, we want methods that provide quantitative information about the nature and magnitude of interactions. Third, we would like the assumptions underlying the method(s) to be as far removed as possible from the detection and description of the interactions of interest. This chapter presents a non-parametric hierarchical Bayesian model designed to pursue these goals. Simulations indicate that the Bayesian model has statistical power equal to that of previously employed frequentist methods (e.g., Thomas, 2001a, 2001b), and so is as reliable a detector of interactions as is (currently) possible. The Bayesian model also provides more information about the direction and magnitude of interactions and relies on fewer and less directly related assumptions than do analogous frequentist methods. The Bayesian model has a number of additional advantages over these frequentist methods.

3.1. Tests of Marginal Response Invariance and Report Independence

In previous work, tests of report independence have been carried out via χ^2 tests of independence (of the stimulus dimensions); response frequencies for each stimulus are arranged in a 2x2 contingency table and the familiar χ^2 formula is used to calculate a test statistic. Similarly, tests of marginal response invariance have been carried out via a z test (see, e.g., Thomas, 2001a). In the present work, analogous tests are carried out with a hierarchical non-parametric Bayesian statistical model. As described below in detail, the Bayesian model provides us with posterior distributions of estimates of identification confusion probabilities. These estimates can be transformed to obtain measures of marginal response invariance and report independence, the posterior distributions of which provide

information regarding the direction and magnitude of interactions as well as quantification of the uncertainty of observed values of measures of interaction.

By taking the difference of the terms on each side of equation 1 (i.e., the numerator in equation 11), we obtain a statistic for testing whether or not marginal response invariance holds (in this case for level i on dimension A ; we do so analogously for each level on dimension B). We call this statistic Δ as it indicates a difference of marginals.

$$(4) \quad \begin{aligned} \Delta &= p(a_i|A_i, B_1) - p(a_i|A_i, B_2) \\ &= \left[p(a_i, b_1|A_i, B_1) + p(a_i, b_2|A_i, B_1) \right] - \left[p(a_i, b_1|A_i, B_2) + p(a_i, b_2|A_i, B_2) \right] \end{aligned}$$

Similarly, we take the difference between the terms in equation 3 to obtain a useful statistic for testing report independence for each stimulus (i.e., each row in a confusion matrix). We call this statistic π as it indicates a test of (in part) *p*erceptual *i*ndependence:

$$(5) \quad \begin{aligned} \pi &= p(a_i, b_j|A_i, B_j) - \left[p(a_i|A_i, B_j) \times p(b_j|A_i, B_j) \right] \\ &= p(a_i, b_j|A_i, B_j) - \\ &\quad \left(\left[p(a_i, b_1|A_i, B_j) + p(a_i, b_2|A_i, B_j) \right] \times \left[p(a_1, b_j|A_i, B_j) + p(a_2, b_j|A_i, B_j) \right] \right) \end{aligned}$$

The Δ statistic can, in theory, detect an interaction (i.e., a failure of marginal response invariance) and, because it is signed, indicate the nature and magnitude of the interaction. For example, if marginal response invariance were to fail due to an increased probability of responding a_1 (e.g., ‘labial’) when presented with the A_1, B_1 stimulus (e.g., ‘labial, voiceless’ [p]) than when presented with the A_1, B_2 stimulus (‘labial, voiced’ [b]), Δ (as defined in equation 4) would be positive and would reflect the magnitude of the difference in response probabilities. The z statistic employed by Thomas (2001a) behaves similarly.

On the other hand, the χ^2 test of (report) independence can only serve to detect an interaction. Because χ^2 values are by definition positive, information about the nature of the interaction is lost, and because of the formula for the χ^2 statistic, the relationships between the magnitude of failure of report independence (i.e., of the difference between the joint probability and the product of marginals in equation 5) and the magnitude of χ^2 statistic is opaque. The π statistic defined in equation 5 preserves this information. However, it is important to note that, while the behavior of the z and

χ^2 statistics are well-understood, the same cannot be said for the ‘raw’ differences that constitute Δ and π . We show below (in Appendix A) that, in the context of the Bayesian model employed here, Δ and π have equivalent statistical power as z and χ^2 (i.e., are as likely to accurately detect an underlying perceptual or decisional interaction). Only Δ and π will be presented and discussed here.

3.2. The Bayesian model

A hierarchical Bayesian multinomial-Dirichlet model provides a number of distinct advantages over frequentist null-hypothesis significance tests of marginal response invariance and report independence. First, the hierarchical structure¹ provides both group level and individual subject level information about the distribution of correct identifications and incorrect confusions (and, by implication, marginal response invariance and report independence); previous work in the GRT framework has focused (statistically) almost exclusively on the individual-subject level. Second, the Bayesian model proposed here avoids some of the limitations of null-hypothesis significance testing. The Bayesian model described below provides a distribution of estimates of identification-confusion probabilities that are reasonably probable given a well-specified pair of assumptions, namely that the identification-confusion counts are multinomial random variables (equation 7, below) and that the underlying probabilities driving the observed identification-confusion counts of each subjects are jointly distributed as Dirichlet random variables tied together by group level (modeled) identification-confusion counts (equation 9, below). By way of contrast, frequentist tests of marginal response invariance and report independence are models of these constructs, and so rely on strong assumptions about the absence of interactions (i.e., null hypotheses regarding the value of test statistics).

The third advantage is that a Bayesian model enables straightforward quantitative prediction of future data, given the data observed. Fourth, and again related, the Bayesian model employed here allows us to investigate both variances in marginal distributions of test statistics (i.e., uncertainty with regard to individual tests of marginal response invariance, akin to frequentist confidence intervals) *and* covariances in joint (posterior) distributions of these statistics.² Fifth, and finally, a

¹ which is not inherently Bayesian

² It seems likely that methods to obtain covariance information exist or are possible with the frequentist statistics that have been used in previous GRT work, but to the best of my knowledge, these tools have not been developed or applied. In any case, the covariances are a natural byproduct of the Bayesian analysis.

Bayesian model explicitly encodes prior information. Although this is of limited utility in the present case, due to a general paucity of relevant previous work on which to base a model of prior knowledge, the results of the analyses reported here can serve as a quantitative model of prior knowledge in future experimentation and analyses. The Bayesian model used here is given in general form in equation 6:

$$(6) \quad p(\theta, \alpha | data) \propto p(data | \theta) p(\theta | \alpha) p(\alpha)$$

Here, $p(\theta, \alpha | data)$ is the probability of the individual subject level estimates of identification confusion probabilities θ and the group level estimates of identification confusion counts α , $p(data | \theta)$ is the likelihood of the data given the θ probability estimates, $p(\theta | \alpha)$ is the likelihood of the θ probability estimates given the α count estimates, and $p(\alpha)$ is the (hyper) prior probability of the α count estimates. The symbol \propto means ‘proportional to’, so the posterior probability of θ and α is proportional to the product of the likelihood, the prior, and the hyper-prior.

The likelihood of the data is defined as follows. The frequencies of the responses to a given stimulus are distributed as multinomial random variables. Thus, for a single subject and a single stimulus, the likelihood of the data for a given set of θ values is given by equation 7, where $N!$ is the factorial of the number of presentations of the stimulus, and $n_j!$ is the factorial of the frequency of the j^{th} response, and $0 \leq \theta \leq 1$:

$$(7) \quad \begin{aligned} p(data | \theta) &= p(\mathbf{n}_j | \theta) \\ &= \frac{N!}{n_1! n_2! n_3! n_4!} \prod_j \theta_j^{n_j} \\ &\propto \prod_j \theta_j^{n_j} \end{aligned}$$

The ratio of factorials is a normalizing constant for any given data set, and so can be ignored for the purposes of illustration. The likelihood of a subject’s entire data set is the product of the likelihoods of each of that subject’s stimuli, and the likelihood of multiple subjects’ data sets is the product (of products) across subjects. This is given in equation 8 for each stimulus i , response j , and subject k :

$$(8) \quad p(\text{data}|\theta) = p(\mathbf{n}_{ijk}|\theta) \\ \propto \prod_{i,j,k} \theta_{ijk}^{n_{ijk}}$$

There is an underlying probability corresponding to each combination of subject, stimulus, and response, so there are as many θ estimates as there are observed response frequencies (i.e., confusion matrix cells). These are constrained somewhat, and the group level behavior is modeled, by defining the likelihood of the individual level θ probability estimates given group level α count estimates as a Dirichlet distribution ($\alpha > 0$):

$$(9) \quad p(\theta|\alpha) \propto \prod_{i,j,k} \theta_{ijk}^{\alpha_{ij}}$$

Finally, the prior distribution of the α count estimates is defined as a Gamma distribution with rate and shape parameters a and b :

$$(10) \quad p(\alpha) \sim \text{Gamma}(a, b)$$

In each analysis, a and b were set equal to 3 and .02, respectively, giving a distribution of α values with a mean of 150 and standard deviation of approximately 86. This Gamma distribution is illustrated in Figure 3.1.

Estimation of the posterior distributions on θ and α were carried out in BRUGS, a software package linking the BUGS (Bayesian Updating via Gibbs Sampling; see, e.g., <http://mathstat.helsinki.fi/openbugs/>) software to the statistical software package R (Team, 2009). The BRUGS code used is given in Appendix B. BRUGS samples from the posterior distribution of θ and α (the left side of equation 6) according to the model defined in equations 8, 9, and 10. Samples of θ and α values are output in vectors, which are then analyzed to investigate overall model fit, marginal response invariance, and report independence.³ As indicated in the BRUGS code, 3 chains (i.e., separate sets of posterior samples) were acquired in order to check for convergence of the sampling algorithm to the true posterior distribution of θ and α . The first 10,000 steps of each chain were discarded as ‘burn-in’, and each chain was thinned such that only every 200th step was taken (to minimize

³ Micro- and macro-analytic signal detection parameters can also be calculated from these vectors, though this not done here.

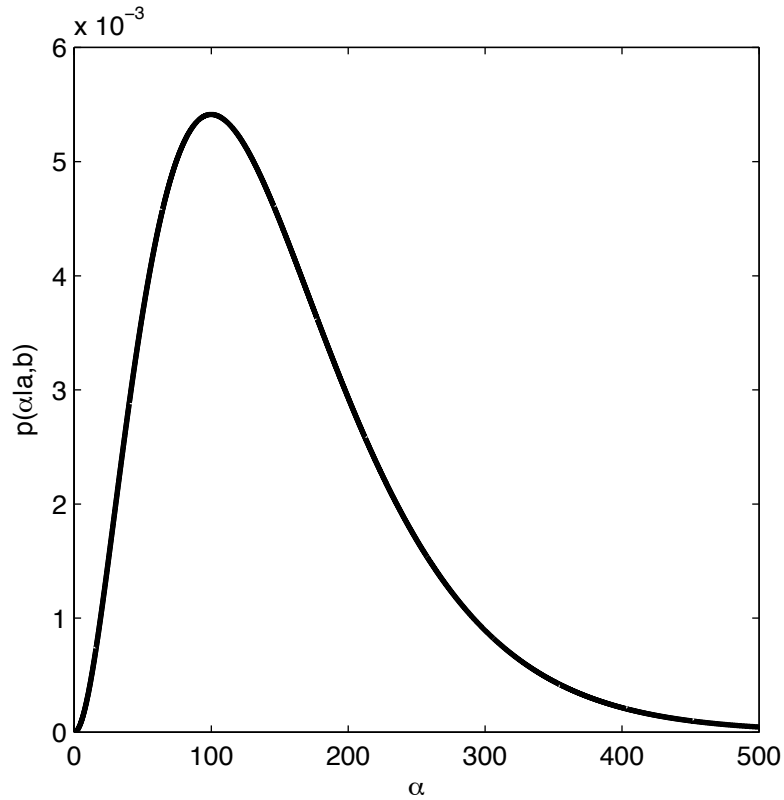


FIGURE 3.1. The Gamma distribution used as the (hyper) prior distribution on the α count estimates. Gamma distribution parameters $a = 3$, $b = 0.02$.

autocorrelation of posterior samples) such that the output chains to be analyzed each consisted of 1000 samples (i.e., each chain had length 1000). Stability and convergence were checked via the Gelman-Rubin statistic, roughly an F -like statistic that compares the variance within and between chains. Given sufficiently low autocorrelation and good evidence of convergence, the three chains are concatenated to form a single chain (of length 3000) corresponding to each θ and α in the model.

Overall fit of the model to each subject's data is evaluated by comparing observed identification-confusion proportions to the posterior distribution(s) of θ values. Marginal response invariance and report independence are evaluated by calculating, for each step in the chains, the values of Δ and π (as defined in equations 4 and 5, respectively), and investigating the posterior distributions of same.

3.3. Relationships between known GRT models, Δ , and π

This section is designed to make explicit the relationships between known (Gaussian) GRT models and the Bayesian analyses of marginal response invariance and report independence. It is also intended to familiarize the reader with the format to be used in the presentation of analyses in later chapters, and to discuss the kinds of inferences that can be drawn from the analyses. Data was simulated from six Gaussian GRT models and analyzed via the model described above in Section 3.2.

Figure 3.2 shows a model (left panel) that exhibits all three forms of independence defined in GRT – perceptual independence, perceptual separability, and decisional separability. The results of analyses of marginal response invariance on the place of articulation dimension are shown in the top right panel; the x -axis indicates Δ_{labial} values, and the y -axis indicates $\Delta_{alveolar}$ values. In this case, the posterior distributions of Δ_{labial} and $\Delta_{alveolar}$ are both clustered tightly around zero, indicating (correctly) that marginal response invariance should be produced by the model at both levels of the place dimension⁴ and that the data are consistent with only small deviations (of Δ) from zero.

The results of analyses of report independence are shown in the bottom right panel; the x -axis indicates π_p values, the y -axis π_b values. The posterior distributions of both π_p and π_b are also clustered tightly around zero, indicating (correctly) that report independence holds. Note, though, that there is a shift toward positive values for each. This is due to the shifted (relative to the perceptual distribution means) location of the decision bounds (left panel).

Figure 3.3 shows two GRT models that produce similar patterns of failure of report independence. The two panels on the left show the models used to simulate data. The model shown in the top left panel exhibits the presence of perceptual independence and failures of decisional separability that increase the size of the response region for ‘p’ and decrease the size of the response region for ‘b’.⁵ Relative to a case in which decisional separability holds, more of the [p] perceptual distribution falls in the correct response region than in the adjacent partially correct regions (i.e., the ‘labial’-‘voiced’, or ‘b’, region in the top left and the ‘alveolar’-‘voiceless’, or ‘t’, region in the bottom

⁴ Recall that Ashby and Townsend (1986) show that perceptual and decisional separability together imply marginal response invariance.

⁵ Similar changes also occur in the response regions for ‘t’ and ‘d’; for the sake of simplicity, we focus only on ‘p’ and ‘b’ here.

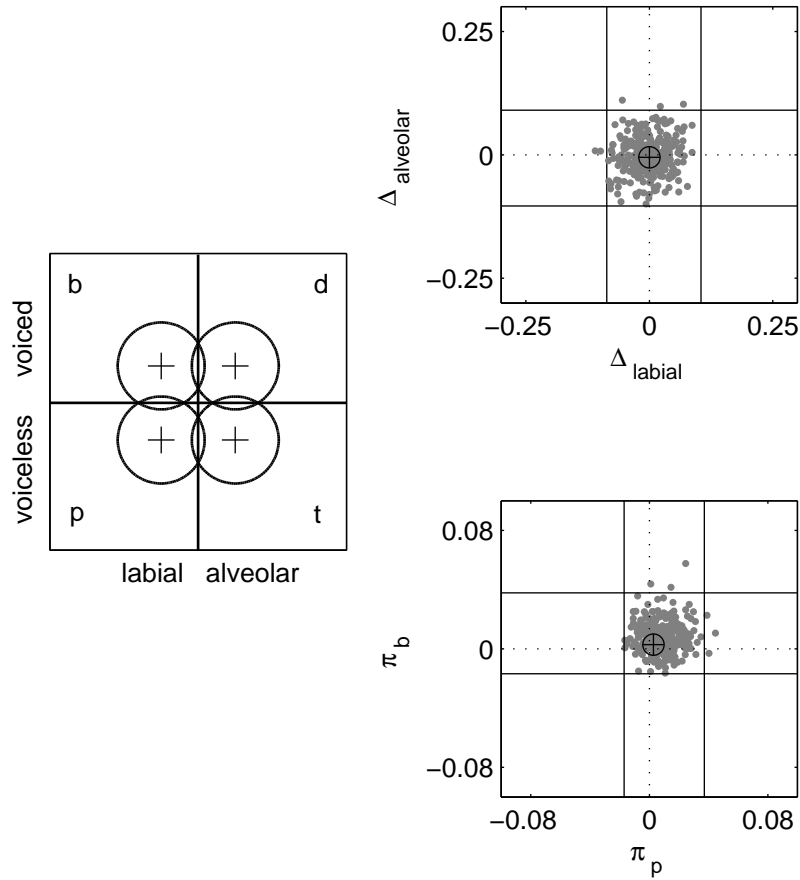


FIGURE 3.2. A Gaussian GRT model exhibiting perceptual independence, perceptual separability, and decisional separability (left panel) and a subset (1/10) of the posterior samples of Δ (gray cloud, top right panel) and π (gray cloud, bottom right panel). Solid lines (in the two right panels) indicate limits on 99% highest probability density regions (i.e., regions in which 99% of the posterior distribution resides). Circled crosses indicate values of Δ (top right panel) and π (bottom right panel) calculated from the simulated data.

right). Thus, because π is defined as the probability of the correct response minus the product of the probabilities of the correct marginal responses (see equation 5), we expect positive values of π_p . On the other hand, relatively less of the [b] perceptual distribution falls in the correct response region than in the adjacent partially correct regions (i.e., the ‘labial’-‘voiceless’, or ‘p’, region in the

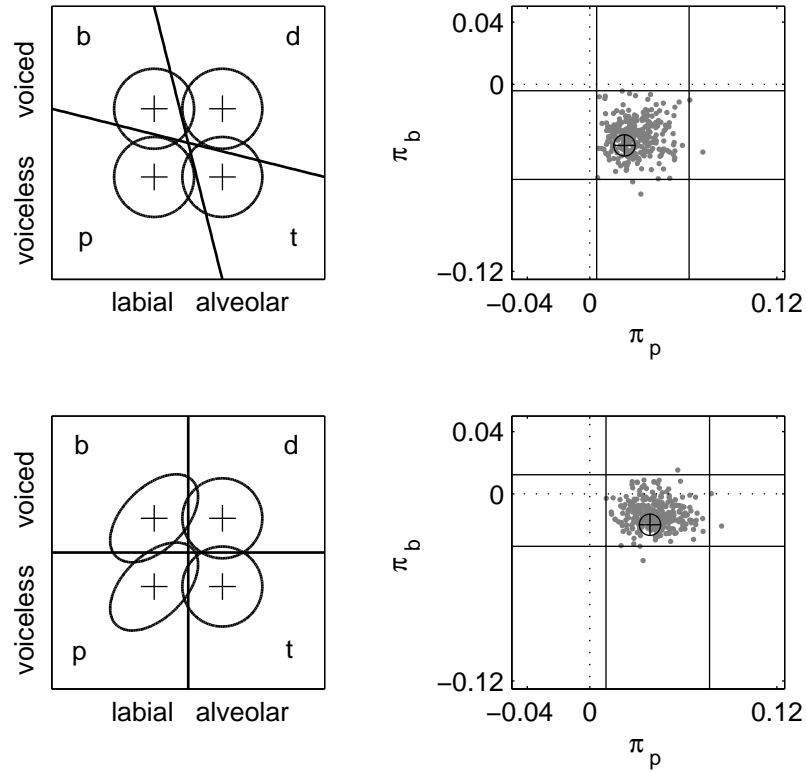


FIGURE 3.3. Gaussian GRT models exhibiting failure of decisional separability (top left panel) and failure of perceptual independence (bottom left panel) and a subset (1/10) of the posterior samples of π (gray clouds, right panels) resulting from analysis of simulated data. Solid lines (in the two right panels) indicate limits on 99% highest probability density regions (i.e., regions in which 99% of the posterior distribution resides). Circled crosses indicate values of π (right panels) calculated from the simulated data.

bottom left and the ‘alveolar’-‘voiced’, or ‘d’, region in the top right). Hence, we expect negative values of π_b . The top right panel shows exactly this pattern of results for both the observed (circled crosses) and modeled (gray clouds) values.

The model depicted in the bottom left panel exhibits the presence of decisional separability and the failure of perceptual independence (in the [p] and [b] perceptual distributions) such that, again,

relatively more of the [p] perceptual distribution and relatively less of the [b] perceptual distribution fall in the (appropriate) correct response regions than in the adjacent partially correct regions. A similar pattern of results – positive π_p and negative π_b – is expected and observed. The magnitude of the shifts in the posterior distributions is different for the two models, but the general pattern is the same.

This difference in magnitude highlights an important issue in drawing inferences from the results of a Bayesian analysis such as those employed here. One way to draw inferences from a posterior distribution of, e.g., π values is to define a ‘highest probability density’ region and determine whether or not a particular value of interest (e.g., zero) is contained within this region. For example, in the presence of perceptual independence and decisional separability, we expect a value of π near zero. If zero is excluded from the highest probability density region of choice, we can conclude, with some degree of confidence, that report independence has failed (i.e., that the observed value of π is [relatively] inconsistent with a true underlying value of zero). In the ‘failure of decisional separability’ case illustrated in the top two panels of Figure 3.3, the 99% highest probability density regions for π_p and π_b exclude zero, whereas in the ‘failure of perceptual independence’ case illustrated in the bottom two panels, zero is only excluded from the 99% highest probability density region of π_p , but not that of π_b .

With regard to π_b , in this case, the discrete inference is, in some sense, incorrect, as it has failed to detect a true underlying interaction. Nonetheless, the analysis provides valuable information about this interaction. Although zero is within the 99% highest probability density region of the posterior distribution of π_b , the distribution is shifted substantially in the expected negative direction. Most of the posterior distribution consists of negative values of π_b , which provides us with information about the state of report independence in the data. Given the imbalance in favor of negative π_b values, it is reasonable to draw inferences about perceptual independence and decisional separability in the underlying model, although our degree of confidence in such inferences is not – and should not be – as high here as it is in the other three cases (in which zero is excluded from the 99% highest probability density regions).

Figure 3.4 depicts two GRT models that produce similar patterns of failure of marginal response invariance. The two panels on the left show the models from which simulated data were produced. The model shown in the top left panel exhibits failure of decisional separability and the presence

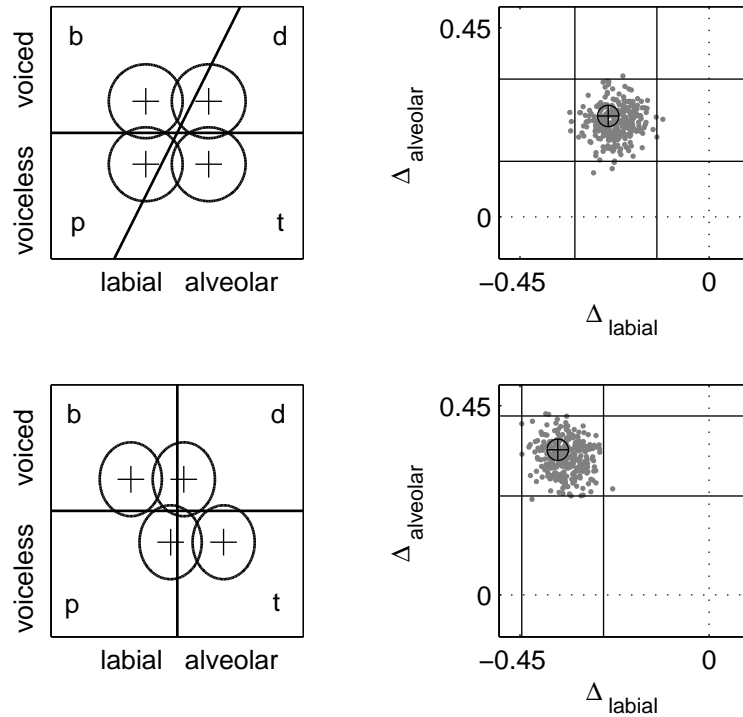


FIGURE 3.4. Gaussian GRT models exhibiting failure of decisional separability (top left panel) and failure of perceptual separability (bottom left panel) and a subset (1/10) of the posterior samples of Δ (gray clouds, right panels) resulting from analysis of simulated data. Solid lines (in the two right panels) indicate limits on 99% highest probability density regions (i.e., regions in which 99% of the posterior distribution resides). Circled crosses indicate values of Δ (right panels) calculated from the simulated data.

of perceptual separability. In this model, responses to place of articulation depend on the voicing dimension such that the vertical (i.e., ‘place’) decision bound is ‘tilted’ and shifted to produce more ‘labial’ responses in the voiced region of perceptual space and more ‘alveolar’ responses in the voiceless region. Thus, there is a greater probability of responding ‘labial’ when presented with [b] than with [p] because, relative to the [p] distribution, more of the [b] distribution falls on the ‘labial’ side of the vertical decision bound, and there is a greater probability of responding ‘alveolar’ when

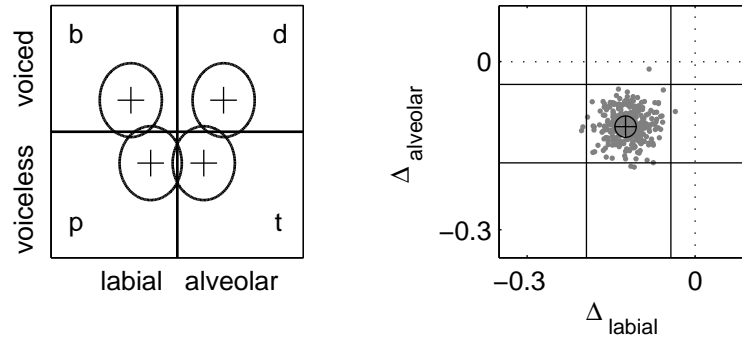


FIGURE 3.5. A Gaussian GRT model exhibiting failure of perceptual separability (left panel) and a subset (1/10) of the posterior samples of Δ (gray clouds, right panel) resulting from analysis of simulated data. Solid lines (in the right panel) indicate limits on 99% highest probability density regions (i.e., regions in which 99% of the posterior distribution resides). The circled cross indicates values of Δ (right panels) calculated from the simulated data.

presented with [t] than with [d] because, relative to the [d] distribution, more of the [t] distribution falls on the ‘alveolar’ side of the bound. As expected (given the definition of Δ in equation 4), the analysis of marginal response invariance on the place dimension produces negative values of Δ_{labial} and positive values of Δ_{alveolar} .

A similar pattern holds for the model exhibiting failure of perceptual separability and the presence of decisional separability (bottom left panel). In this model, there is a similar imbalance in the probability of ‘labial’ responses when presented with [p] and [b] and in the probability of ‘alveolar’

responses when presented with [t] and [d], though here this is due to shifts in the means of the distributions, not changes in the location and orientation of the vertical decision bound. In this case, the voiced consonants [b] and [d] produce percepts more toward the labial side of perceptual space, whereas the voiceless consonants [p] and [t] produce more alveolar percepts. The results of the analysis are very similar (bottom right panel). Again, as expected, we see negative values of Δ_{labial} and positive values of $\Delta_{alveolar}$. The magnitude of the shift in the posterior distribution of Δ values is different in the two models illustrated here, but this is an artifact of the particular parameters chosen to simulate the data. The key point is that, while the underlying interactions in the two models are distinct, the ‘observed’ and modeled values of Δ are very similar.

The failure of perceptual separability depicted in the bottom left panel of Figure 3.4 is often called ‘mean-shift integrality’, as perceptual separability has failed (i.e., the stimuli are ‘perceptually integral’) due to a shift in the means of the perceptual distributions. Perceptual separability may also fail if the salience of a given dimension varies as a function of the other dimension. Figure 3.5 illustrates an example of this. The left panel shows a model in which the place contrast between labial and alveolar segments is less salient for voiceless [p] and [t] than it is for voiced [b] and [d]. In this case, the probability of correct place of articulation responses is greater for both voiced stimuli, so the values of Δ should be negative at both levels of the place dimension. This is shown in the right panel (cf. the patterns shown in Figure 3.4).

No failure of decisional separability can produce this pattern. If a failure of decisional separability shifts response bias toward, say, ‘labial’ responses in the voiced region of perceptual space (as in the top left panel of Figure 3.4), it necessarily shifts response bias *away* from ‘alveolar’ responses. The model illustrated in Figure 3.5 produces an increase in (correct) ‘labial’ *and* ‘alveolar’ responses in the voiced region of perceptual space. Note, too, that if salience were lower for voiced consonants than for voiceless, this would produce positive Δ_{labial} and $\Delta_{alveolar}$ values (because of the order of the terms in the definition of Δ in equation 4). The general point is that ‘mean-shift’ failures of perceptual separability and analogous failures of decisional separability should produce opposite signed shifts in Δ at the two levels on the dimension in question, whereas a failure of perceptual separability due to a change in salience of one dimension across levels of the other should produce shifts in Δ with the same sign at each level.

3.4. Summary

The hierarchical, Bayesian multinomial-Dirichlet model described in equations 6 through 10 may be employed to draw inferences about perceptual and decisional interactions from data. This approach has a number of advantages over frequentist methods employed elsewhere. Chief among these advantages are the fact that the hierarchical model described here relies only on general assumptions about identification-confusion frequencies and explicitly models group-level properties. Previously employed frequentist methods rely on strong assumptions about the measures of interaction of interest and do not explicitly address groups of subjects. In conjunction with the statistics Δ and π , the model also provides information about the direction and magnitude of failures of marginal response invariance and report independence that is often ignored in analogous frequentist tests. Finally, the model enables straightforward visual representation of, on the one hand, the output of the Bayesian analyses, and, on the other, the connections between these analyses and the assumed underlying GRT models.

Place of articulation and voicing in onset position

Experiment one is an investigation of perceptual independence, perceptual separability, and decisional separability between place of articulation and voicing in onset (i.e., syllable initial) stop consonants [p], [b], [t], and [d]. These consonants were chosen in part because previous work on dimensional interactions in speech perception has probed these dimensions with the same consonants, in part because they are one set (among many in English) that mirror the structure of a simple GRT model (i.e., two levels on each of two dimensions). Previous work investigating the integration of various acoustic-phonetic cues to voicing and place (e.g., VOT, formant frequency at voicing onset) in categorization tasks provides some evidence that place and voicing are independent (e.g., Oden & Massaro, 1978). On the other hand, there is also evidence of interactions between place and voicing (e.g., Eimas et al., 1978; Sawusch & Pisoni, 1974). However, within- and between-category perceptual levels are conflated in Oden and Massaro (1978), perceptual and decisional levels are conflated in Eimas et al. (1978), and strong assumptions about the relevant acoustic cues to voicing and place limit the generality of these findings. The application of GRT and naturally produced stimuli addresses both of these concerns.

4.1. Methods

4.1.1. Stimuli. In order to avoid strong assumptions about the relevant acoustic-phonetic dimensions, naturally produced nonsense syllables were used as stimuli. Four tokens of each stimulus type – [pa], [ba], [ta], and [da] – were produced by the author (at the time of recording, a 34 year old midwestern, mail phonetician). Multiple acoustic measurements (given in Appendix C) were taken and extensive pilot experimentation was carried out to ensure both that no particular token was overly acoustically distinct¹ and that the stimuli were within the normal range of values for these consonants. The stimuli for both experiments were recorded during a single session in a quiet room

¹ Early pilot work with one token per category made it clear that very small, and otherwise irrelevant, differences between stimuli could serve as reliable cues to correct responses.

via an Electrovoice RE50 microphone and a Marantz PMD560 solid-state digital recorder at 44.1 kHz sampling rate with 16-bit depth.

Naturally produced (i.e., not (re)synthesized) tokens are very acoustically distinct, however, and identification data with very high accuracy is not particularly informative with regard to perceptual or decisional independence (or interactions). Thus, stimuli were embedded in ‘speech-shaped’ noise (i.e., white noise filtered such that higher frequencies had relatively lower amplitude than lower frequencies).

4.1.2. Procedure. Each participant was seated in a double-walled sound attenuating booth with four ‘cubicle’ partitions. One, two, or three participants could run simultaneously, each in front of his or her own computer terminal. Stimuli were downsampled and presented at -3 dB signal-to-noise ratio at approximately 60 dB SPL via Tucker-Davis-Technologies Real-Time processor (TDT RP2.1; sampling rate 24,414 Hz), programmable attenuator (TDT PA5), headphone buffer (TDT HB6), and Sennheiser HD250 II Linear headphones. Before the first session (familiarization and training), participants read a written instruction sheet, were given verbal instructions, and were prompted for questions about the procedure. Sessions lasted from one to four experimental blocks. Experimental blocks lasted approximately 25 minutes.

Each experimental block began with brief written instructions reminding participants to respond as accurately and as quickly as possible and providing explicit guessing advice for trials on which the participant was uncertain of the stimulus identity.² After the instructions were cleared from the screen, four ‘buttons’ corresponding to the buttons on a hand-held button box became visible. On the on-screen buttons the letters ‘p’, ‘b’, ‘t’, and ‘d’ appeared in black text. Button-response assignments were randomly assigned for each block with the constraint that the basic dimensional structure was always maintained (e.g., ‘p’ and ‘t’ always appeared as neighbors on a single dimension, never on opposite corners).

Each trial consisted of the following steps: (1) A visual signal (the word ‘listen’) presented on the computer monitor; (2) half a second of silence; (3) stimulus presentation; (4) response; (5) feedback; and (6) 1 second of silence. Responses were collected via a button box with buttons arranged

² The data analyzed here are a subset of the total data collected. Each subject completed ten blocks, two in each of five stimulus presentation base rate conditions. Only the equal base rate condition (i.e., the condition in which each stimulus was presented equally frequently) will be analyzed and discussed here.

to correspond to the structure of the stimulus space (i.e., two levels on each of two dimensions). Feedback was given visually via color-coded (green for correct, red for incorrect) text above and on the on-screen buttons. Either the word ‘Correct’ or the word ‘Incorrect’ appeared along with brief descriptions of the presented stimulus and the response chosen. The feedback text disappeared and the button text color was reset to black before each successive trial.

Each participant received two short (approximately 15 minutes) and two regular length blocks to familiarize them with the stimuli and ensure that performance was consistently above chance. The data analyzed here consists of 800 trials completed in two blocks of 400 trials each. Participants were paid \$6/hour with a \$4/hour bonus for completion of the experiment. The participant with the highest accuracy received a \$20 bonus, as did the participant with the fastest overall mean response time.

4.1.3. Subjects. Eight adults (two male, six female) were recruited from the university community. The average age of participants was 21 (19-23). All were native speakers of English with 4.5 (1.5-7) years of second language study on average. All but one were right handed, and all but one were from the Midwest (the other was from the West). All participants were screened to ensure normal hearing.

4.1.4. Analysis. Data analysis was carried out as described above in Section 3 of Chapter 2. To recap briefly, the data were analyzed via a hierarchical Bayesian multinomial-Dirichlet model, from which 3000 samples from the posterior distribution of θ (estimated individual subject identification-confusion probabilities) and α (estimated group-level identification confusion counts) values were drawn. Overall model fit for each subject was evaluated by comparing observed identification-confusion proportions to the modeled probabilities (i.e., θ values). Marginal response invariance and report independence were tested by calculating Δ and π for each sample from the posterior distribution of θ and (normalized) α values in order to consider the range of values for these statistics that could credibly underlie the data.

4.2. Results

4.2.1. Model checking. An important first step in the analysis is model checking. If the model does not correspond reasonably closely to the observed data, further inferences regarding marginal

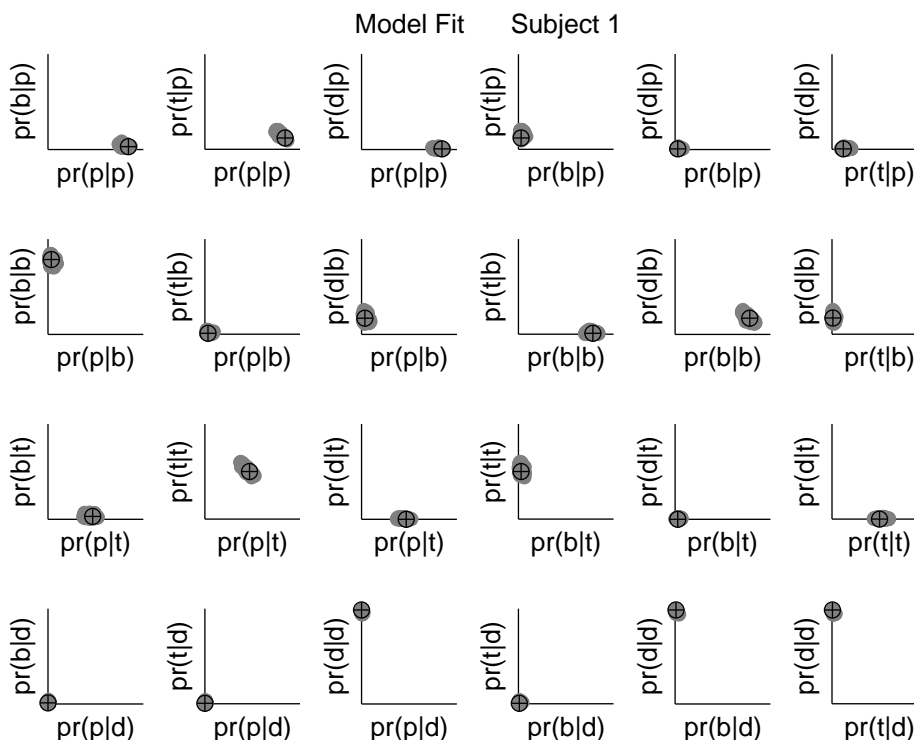


FIGURE 4.1. Scatter plots showing pairs of modeled identification-confusion probabilities (i.e., θ values; gray clouds) and observed identification-confusion proportions (circled black crosses) for subject one.

response invariance and report independence would be questionable. Recall that the multinomial-Dirichlet model described in Chapter 3 has as many degrees of freedom as do the data; there are θ values and response frequencies for each combination of subject, stimulus, and response. Thus, the model should fit the data extremely well.³ In fact, this is the case. Figure 4.1 shows modeled identification-confusion probabilities (θ values) and observed identification-confusion proportions for subject one. Other subjects' plots are substantially identical, differing only in minor details. Note that the observed proportions (indicated by black circled crosses) fall, without exception, well within the range of modeled probabilities (gray clouds), indicating a very good fit of the model to the data.

³ The group-level α counts constrain the individual-level θ probabilities to some degree, so it is logically possible for the model to fail to fit the data well.

The goodness of fit of the model indicates that we can safely draw inferences from analyses of marginal response invariance and report independence statistics.

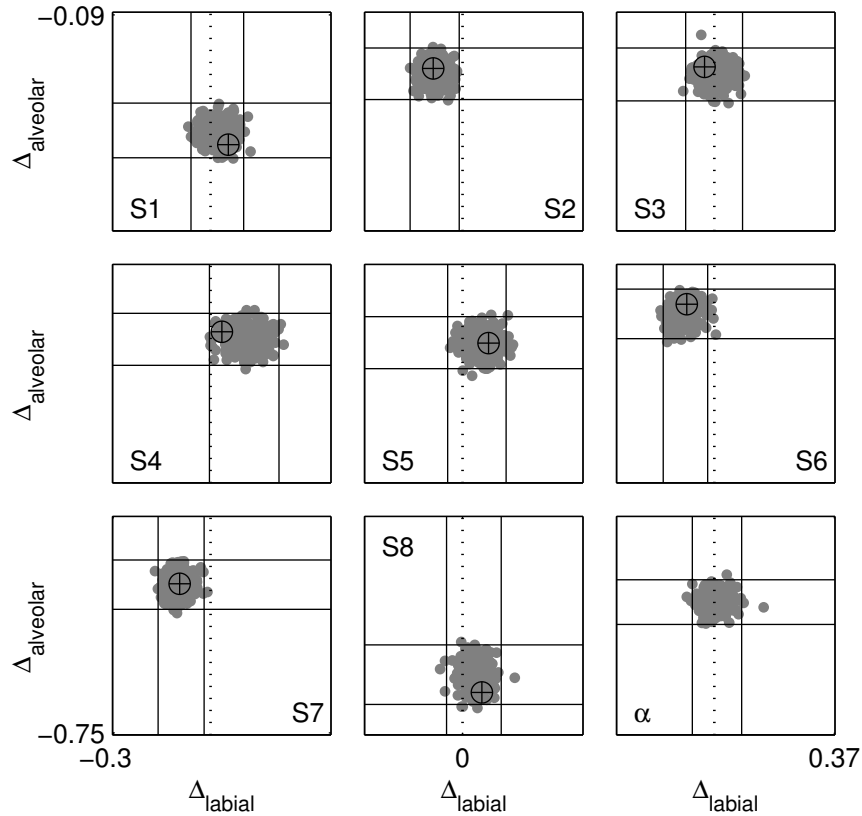


FIGURE 4.2. Scatter plots showing Δ_{labial} and $\Delta_{alveolar}$ (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of sampled posterior Δ values (i.e., transformed posterior θ values); observed statistics are indicated by black circled crosses. The x -axes indicate values of Δ_{labial} (i.e., the probability of responding ‘labial’ when presented with [p] minus the probability of responding ‘labial’ when presented with [b]); the y -axes indicate values of $\Delta_{alveolar}$ (i.e., the probability of responding ‘alveolar’ when presented with [t] minus the probability of responding ‘alveolar’ when presented with [d]). The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

4.2.2. Analysis of marginal response invariance. Figure 4.2 shows the results of the analysis of marginal response invariance for the place of articulation dimension (i.e., Δ_{labial} and $\Delta_{alveolar}$). Note that, though Δ can lie anywhere on the interval $[-1, 1]$, the ranges on the y - and x -axes are restricted and similar in magnitude to one another, but the limits on each axis are different. The analysis indicates that marginal response invariance fails for ‘alveolar’ responses (‘t’ or ‘d’); on the y -axis, all subjects’ highest probability density regions exclude zero, and no modeled (gray clouds) or observed (circled crosses) values lie on or near zero. The analysis also indicates that marginal response invariance holds for all but three subjects for ‘labial’ responses; the upper limits on the highest probability density regions for subjects 2, 6, and 7 fall just below zero on the x -axis, though in each case they are very close.

The (posterior) distribution of Δ values provides information above and beyond the detection of failure of marginal response invariance. For example, although for two subjects – S1 and S3 – the distribution of Δ_{labial} (x -axes) has a roughly equal number of negative and positive values, for three other subjects – S4, S5, and S8 – the distributions have many more positive than negative values (i.e., the distributions sit largely to the right of the vertical zero line in Figure 4.2). By way of contrast, the three subjects whose highest probability density regions for Δ_{labial} exclude zero (S2, S6, and S7) all show the opposite pattern, namely a shift toward negative values of Δ_{labial} . Similarly, although all subjects’ distributions of $\Delta_{alveolar}$ (y -axes) indicate failure of marginal response invariance, individual subjects’ distributions show substantial variation in the magnitude of this failure.

Recall that under the assumptions of the GRT model, perceptual and decisional separability together imply marginal response invariance. Thus, the fact that marginal response invariance tends to hold for ‘labial’ response (i.e., [p] or [b]) suggests that decisional separability holds for place (with respect to voicing) and that perceptual separability holds at the ‘labial’ level on the place dimension. Given this, the failure of marginal response invariance for ‘alveolar’ responses (i.e., [t] or [d]) strongly suggests an underlying model in which a perceptual (but not decisional) separability fails for alveolar stimuli ([t] and [d]) such that [d] is more salient on the place dimension (i.e., more distinct from [b], its labial counterpart) than is [t] (from [p]).

The bottom right panel shows the posterior distribution of group level Δ_{labial} and $\Delta_{alveolar}$ values. Recall that the prior distribution on α values was defined as a gamma distribution. This means

that there is no strict upper limit on possible α values; although very large values are improbable (for the shape and rate parameters used here), a gamma random variable can take any positive value. There is no further constraint on each set of sampled posterior α values. Thus, in order to calculate a group level measure of marginal response invariance (and report independence, as presented below), each set of sampled α values were converted to relative frequencies (i.e., proportions), which were then transformed into Δ values according to equation 4.

Not surprisingly, the group level measure of marginal response invariance closely reflects the pattern observed for the individual subjects; marginal response invariance fails for ‘alveolar’ responses and holds for ‘labial’ responses. The failure with respect to ‘alveolar’ responses is moderate, reflecting the fact that fairly extreme values were only observed for two subjects (S1 and S8). The range of values with respect to ‘labial’ responses is fairly well balanced around the zero line, reflecting the fact that both positive and negative values of Δ_{labial} were observed.

Figure 4.3 shows the results of the analysis of marginal response invariance for the voicing dimension. Note, again, that the ranges spanned on the y- and x-axes are similar in magnitude and different in specific values (and from the values spanned in Figure 4.2). Only subject 2’s results indicate a failure of marginal response invariance for ‘voiceless’ responses (‘p’ or ‘t’); all others subjects’ highest probability density regions include zero on the x -axis. The $\Delta_{voiceless}$ distributions of all subjects save one (S8) have many more negative than positive values, indicating a consistent tendency toward more ‘voiceless’ responses to [t] than to [p]. On the other hand, the distributions of Δ_{voiced} for five subjects – S3, S4, S5, S6, and S8 – indicate a failure of marginal response invariance for ‘voiced’ responses (‘b’ or ‘d’), by virtue of zero falling outside of the 99% highest probability density regions. All subjects’ distributions of Δ_{voiced} are shifted toward negative values, indicating a tendency toward a higher probability of responding ‘voiced’ when presented with [d] than when presented with [b].

As with the analysis of the place dimension, the fact that marginal response invariance generally holds for ‘voiceless’ responses and tends to fail for ‘voiced’ responses suggests that the latter are due to failures of perceptual (but not decisional) separability at this level of the voicing dimension such that [d] is more salient (i.e., distinct from [t], its counterpart on the voicing dimension) than is [b] (from [p]). The group level analysis (bottom left panel) again reflects the individual subject level

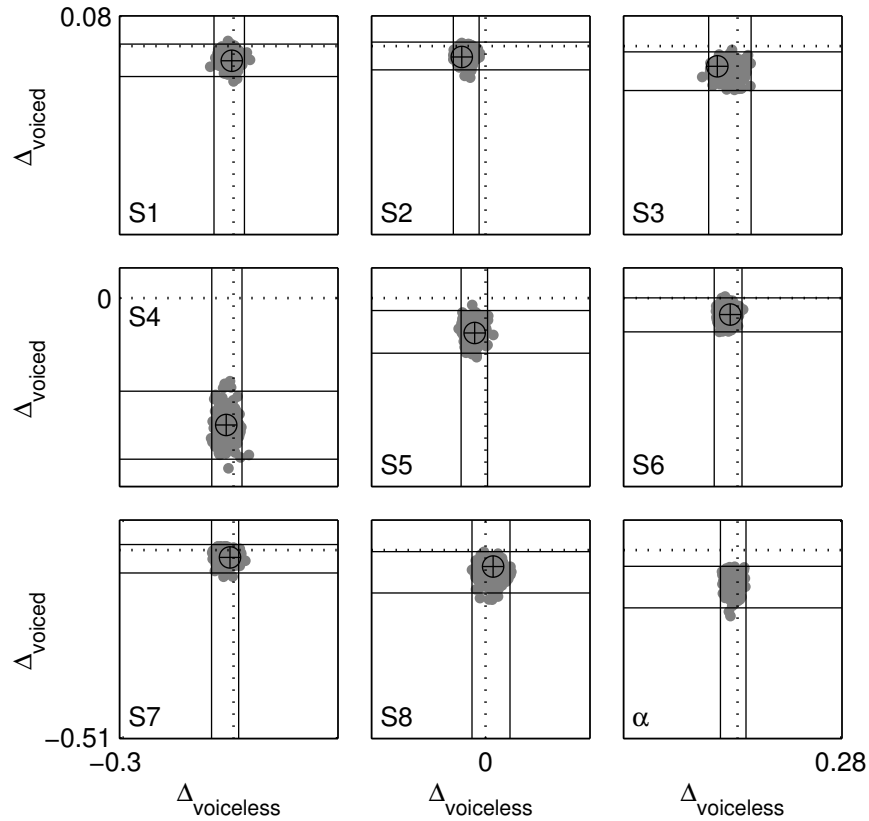


FIGURE 4.3. Scatter plots showing $\Delta_{voiceless}$ and Δ_{voiced} (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the sampled posterior Δ values (i.e., transformed posterior θ values); observed statistics are indicated by black circled crosses. The x -axes indicate values of $\Delta_{voiceless}$ (i.e., the probability of responding ‘voiceless’ when presented with [p] minus the probability of responding ‘voiceless’ when presented with [t]); the y -axes indicate values of Δ_{voiced} (i.e., the probability of responding ‘voiced’ when presented with [b] minus the probability of responding ‘voiced’ when presented with [d]). The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

fairly faithfully in that marginal response invariance holds for ‘voiceless’ responses, with a slight tendency toward negative values, and fails for ‘voiced’ responses.

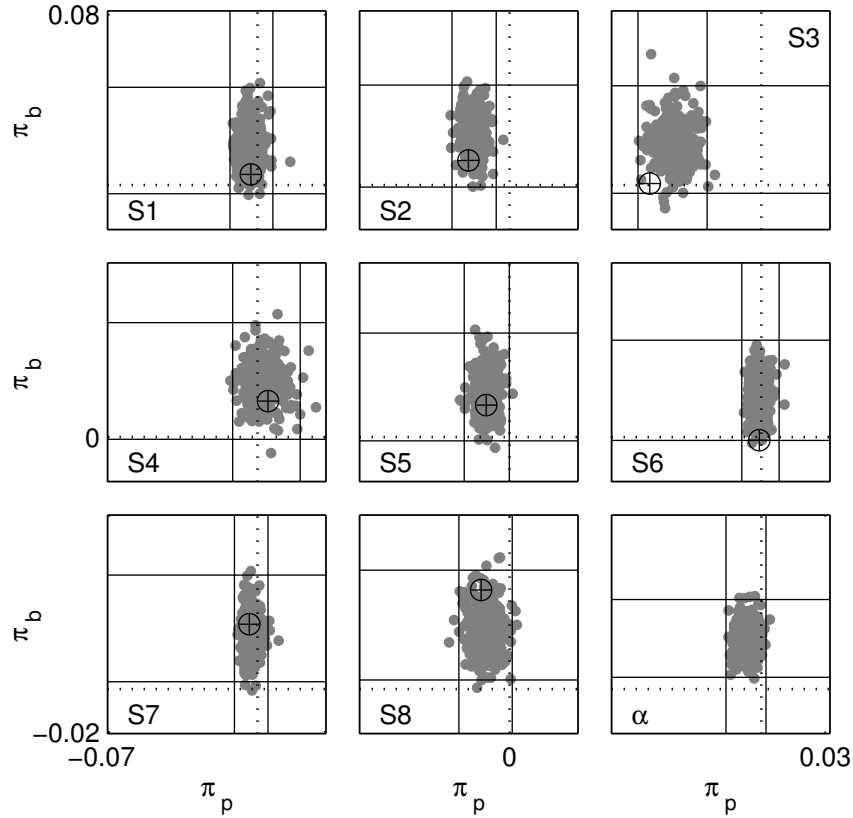


FIGURE 4.4. Scatter plots showing π_p and π_b (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for labial stimuli [p] and [b]. The x -axes indicate the value of π_p ; the y -axes indicate the value of π_b . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

There is considerably smaller variance in the posterior distributions of Δ_{voiced} and $\Delta_{voiceless}$ (see Figure 4.3) than in the posterior distributions of Δ_{labial} and $\Delta_{alveolar}$ (see Figure 4.2). This is likely due to higher accuracy on the voicing dimension than on the place dimension.

4.2.3. Analysis of report independence. Figures 4.4 and 4.5 show the results of the analysis of report independence. Figure 4.4 shows posterior distributions of π_p and π_b , and Figure 4.5 shows posterior distributions of π_t and π_d . Recall that for any given stimulus, the presence of perceptual

independence and decisional separability implies report independence. Recall, too, that because π is a signed statistic, it provides information about the nature of failure of report independence and, by the logical relationship just mentioned, plausible failures of perceptual independence and decisional separability.

Recall that the π statistic is defined (in equation 5) to be the probability of a correct response (i.e., the joint probability of identifying the correct level on each dimension) minus the product of the marginal probabilities of identifying the correct levels on each dimension. As shown in Figure 3.3, a positive (observed or modeled) π statistic indicates failure of perceptual independence, decisional separability, or both, such that (relative to the case with perceptual independence and decisional separability) more of a given perceptual distribution falls in the correct response region than in the adjacent partially correct response regions. On the other hand, a negative π statistic indicates the opposite, namely that perceptual independence, decisional separability, or both, are failing such that relatively less of a given perceptual distribution falls in the correct response region than does in the adjacent partially correct response regions. Note that, though π can, in theory, lie anywhere on the interval $[-1, 1]$, the ranges shown on the y - and x -axes are restricted such that their ranges are similar in magnitude and the limits on each axis are different from one another in Figures 4.4 and 4.5.

The distributions of π_p (x -axes) for all subjects contain roughly equal numbers of positive and negative values, and all subjects' highest probability density regions include zero on the x -axis, indicating that report independence holds for [p]. The distributions of π_b (y -axes) indicate that report independence holds for [b] for all but two subjects – S2 and S5. Only the distribution of S4's π_b show a tendency toward negative values (i.e., relatively less of the perceptual distribution falling the in the correct response region than in the partially correct regions); all other subjects show the opposite tendency. The group level analysis (also) indicates that report independence holds for [p] and [b], though zero is very close to the lower limit of the highest probability density region for π_b and the distribution has far more positive than negative values; zero is squarely in the middle of the group level distribution for π_p .

Figure 4.5 shows a subset of the posterior samples of π_t and π_d . The results for π_t (x -axis) indicate that report independence fails for all but one subject; S1's highest probability density region for π_t excludes zero. All but two subjects' – S3 and S8 – distributions of π_t have far more

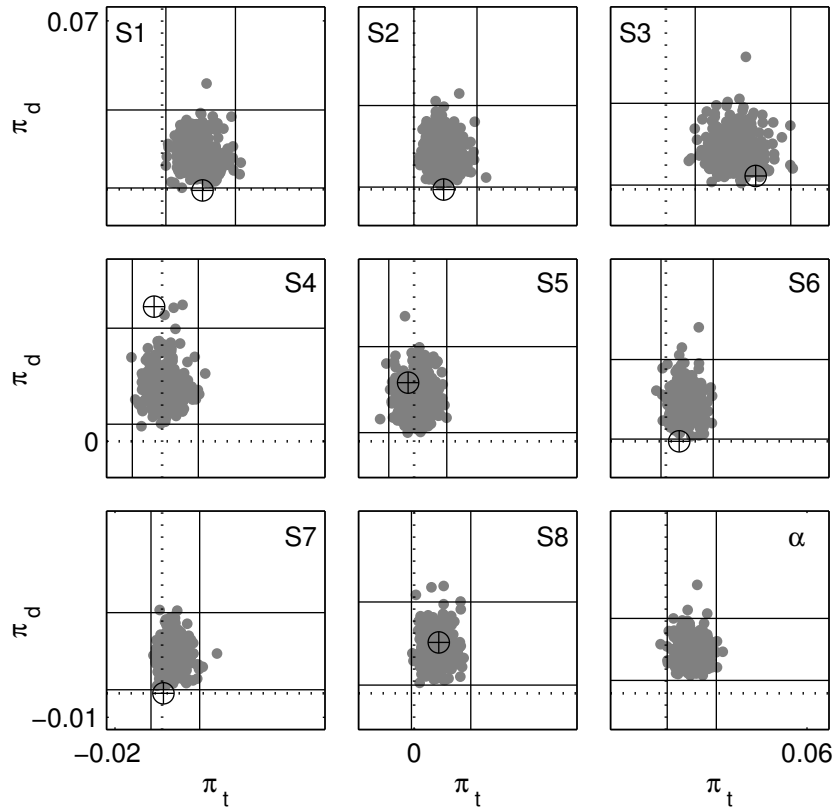


FIGURE 4.5. Scatter plots showing π_t and π_d (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for alveolar stimuli [t] and [d]. The x -axes indicate the value of π_t ; the y -axes indicate the value of π_d . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

positive than negative values, suggesting that relatively more of the [t] perceptual distribution lies in the correct response region than in the neighboring partially correct response regions.

Three subjects' – S4, S6, and S8 – highest probability density regions for π_d exclude zero. All subjects' distributions have substantially more positive than negative values, indicating a relatively greater proportion of the [d] perceptual distribution in the correct response region than the adjoining partially correct regions. Accuracy was very high for [d] for all subjects, though, indicating that most

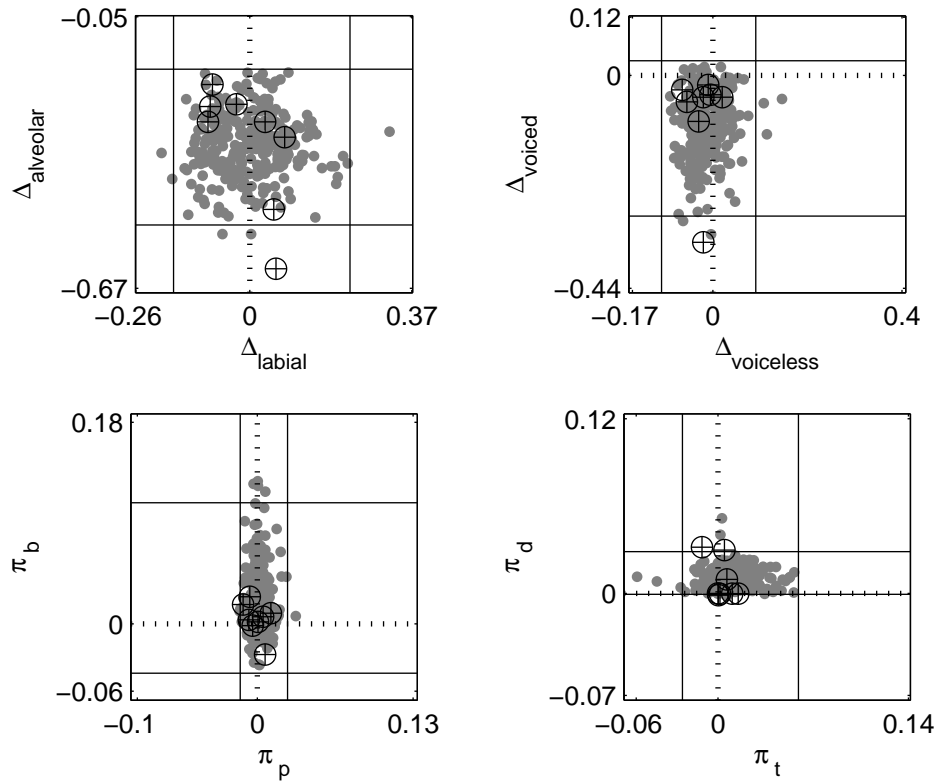


FIGURE 4.6. Scatter plots showing predicted and observed Δ and π statistics. The gray clouds show a subset (1/10) of the predicted values of Δ and π generated from the posterior distribution of the group level α ‘counts’. Black circled crosses indicate observed values of Δ and π . Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

of the [d] perceptual distribution was in the correct response region, which shifts the distribution of π values in a positive direction.

The group level analysis for π_t and π_d suggests that report independence holds for [t] but not for [d], and reflects the general patterns of the individual subject analyses. The modeled values of both π_t and π_d at the group level are far more likely to be positive than negative.

4.2.4. Prediction of future data. One important advantage to employing a Bayesian statistical model is the ease with which quantitative predictions can be made regarding future data. In order to maximize the generality of the predictions, the group level α values were used to generate

predicted identification-confusion probabilities, which were, in turn, used to generate predicted response frequencies. These were then transformed into Δ and π values, shown (with observed values for all eight subjects) in Figure 4.6.

Given the observed identification-confusion data, the demographic properties of the subjects, and the experimental materials and protocol, the Bayesian model predicts that marginal response invariance will hold for ‘labial’, ‘voiceless’, and ‘voiced’ responses, with roughly equal proportions of positive and negative values of Δ_{labial} and $\Delta_{voiceless}$ and a strong tendency toward negative values of Δ_{voiced} . Marginal response invariance is predicted to fail for ‘alveolar’ responses. Report independence should hold for all four consonants, with a narrow range of predicted values on either side of zero for π_p , a wider range of values on either side of zero for π_t . Mostly positive π_b and π_d values are predicted, and the former is predicted to span a much larger range of values.

4.2.5. Summary of results. Table 4.1 shows a summary of the discrete inferences drawn from the analyses of marginal response invariance and report independence (i.e., inferences based on 99% highest probability density region limits) as well as the proportion of correct responses for each subject. Accuracy ranged from 61% to 86%, and bears no obvious relationship to patterns of (failure of) marginal response invariance or report independence. Taken together, the results indicate, on the one hand, that marginal response invariance tends to hold for ‘labial’ and ‘voiceless’ responses and fail for ‘alveolar’ and ‘voiced’ responses, and, on the other, that report independence tends to hold for all four of the stop consonants – [p], [b], [t], and [d] – probed here. While there were a small number of exceptions, this pattern holds for nearly all of the experimental subjects. Group level analyses and prediction of future data are consistent with this pattern as well.

The analyses of marginal response invariance and report independence provide important clues as to the nature of the assumed underlying model of perception and decision making. The fact that marginal response invariance tends to hold for ‘labial’ and ‘voiceless’ responses and the fact that report independence tends to hold for all four stimuli suggests strongly that decisional separability holds for both decision bounds. Furthermore, when marginal response invariance fails for both ‘labial’ and ‘alveolar’ responses (as it does for S2, S5, and S7), it does so in the pattern shown in Figure 3.5, which is to say that the relevant Δ values have the same sign, indicating a change in salience of place across levels of voicing. This pattern cannot be accounted for by a failure of decisional separability. If decisional separability holds for both bounds, then the failures of marginal response

TABLE 4.1. Summary of the results of analyses of marginal response invariance (Δ) and report independence (π). $p(C)$ = proportion correct responses. *lab.* = labial ([p] or [b]); *alv.* = alveolar ([t] or [d]); *-voi* = voiceless ([p] or [t]); *+voi* = voiced ([b] or [d]). \checkmark indicates that marginal response invariance or report independence holds (i.e., that zero is within the 99% highest probability density region) for a given subject for that level and dimension (Δ) or stimulus (π).

Sub.	$p(C)$	$\Delta_{lab.}$	$\Delta_{alv.}$	Δ_{-voi}	Δ_{+voi}	π_p	π_b	π_t	π_d
1	0.78	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
2	0.84				\checkmark	\checkmark		\checkmark	\checkmark
3	0.74	\checkmark		\checkmark		\checkmark	\checkmark	\checkmark	\checkmark
4	0.61	\checkmark		\checkmark		\checkmark	\checkmark	\checkmark	
5	0.78	\checkmark		\checkmark		\checkmark		\checkmark	\checkmark
6	0.87			\checkmark		\checkmark	\checkmark	\checkmark	
7	0.86			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
8	0.68	\checkmark		\checkmark		\checkmark	\checkmark	\checkmark	
α		\checkmark		\checkmark		\checkmark	\checkmark	\checkmark	

invariance and report independence can be taken to indicate failures of perceptual separability and perceptual independence, respectively.

Consider, for example, S1, whose results indicate that marginal response invariance holds for ‘labial’, ‘voiceless’, and ‘voiced’ responses and fails for ‘alveolar’ responses and that report independence only fails for [t]. In this case, it is reasonable to conclude that decisional separability holds, that voicing is perceptually separable from place of articulation, that place is *not* perceptually separable from voicing, and that perceptual independence fails for [t] and holds for [p], [b], and [d]. Similar lines of reasoning lead to similar conclusions for the other subjects and the group level results.

4.3. Discussion

The fact that voicing tends not to be perceptually separable from place is not particularly surprising, as it is well understood that, for example, the amplitude of the stop release burst tends

to be greater in alveolars than labials (see, e.g., Stevens, 1998). For a particular signal-to-noise ratio, then, we should expect the voicing contrast to be more salient in alveolars than labials. Six subjects – all but S1 and S7 – exhibit failures of marginal response invariance for only one level on the voicing dimension, in each case due to negative Δ values (marginal response invariance holds for both voicing levels for S1 and S7), indicating that the alveolar stops were, in fact, more salient than the labial stops.

The fact that place is not perceptually separable from voicing is perhaps a bit more surprising, though a plausible explanation is readily found in the fact that so-called voicing ‘cutback’ in onset voiceless stops destroys much of place information transmitted via formant transitions (see, e.g., Stevens, 1998). Five subjects – S1, S3, S4, S5, and S8 – exhibit a failure of marginal response invariance for ‘alveolar’ responses while marginal response invariance holds for ‘labial’ responses; the other three subjects exhibit failure of marginal response invariance at both levels on the place dimension. Consistent with the ‘cutback’ explanation, in every case, the failures are due to negative Δ values, suggesting a change in salience such that [d] is more distinct from [b] than [t] is from [p].

The apparent failures of perceptual separability can be interpreted (phonologically) as incomplete perceptual generalization of features. On the other hand, the evidence in favor of decisional separability indicates that, with regard to decision making and response bias, place and voicing *do* generalize across one another. The tendency toward presence of perceptual independence suggests that there is no cross-talk between the (cognitive) channels processing voicing and place (e.g., Ashby, 1989), despite the apparent interaction of (speech shaped) masking noise with voicing and place cues resulting in failure of perceptual separability. The fact that perceptual independence holds while perceptual separability fails highlights the importance of this conceptual distinction.

Place of articulation and voicing in coda position

Experiment two is an investigation of perceptual independence, perceptual separability, and decisional separability between place of articulation and voicing in coda (i.e., syllable final) stop consonants [p], [b], [t], and [d]. At an abstract level of analysis, the (phonological) contrasts are the same regardless of higher level (e.g., syllable) structure. In terms of the phonetic implementation of these contrasts, though, it is well known that place and voicing contrasts are different in coda and onset position. Whereas VOT and stop release burst intensity serve as cues for voicing in onset position, they do not in coda position. On the other hand, vowel and consonant closure duration play a key role in cuing voicing distinctions in coda stops; voiced stops are preceded by longer vowels and are themselves shorter in duration than are voiceless stops (Stevens, 1998). Similarly, where formant transitions following consonant release can cue place distinctions in voiced onset stops, and where VOT cutback destroys potentially valuable information regarding place in voiceless onset stops, the formant-based cues to place are present in, and precede, both voiced and voiceless coda stops.

5.1. Methods

5.1.1. Stimuli. As in experiment one, naturally produced nonsense syllables were employed in experiment two. Four tokens of each stimulus type – [ap], [ab], [at], and [ad] – were produced by the author. Multiple acoustic measurements (given in Appendix C) were taken and extensive pilot experimentation was carried out to ensure that, within each category, all tokens were reasonably close together in acoustic space and within a reasonable range of values for these consonants. The stimuli were recorded during the same session during which the recordings for experiment one were made. Speech-shaped noise maskers were produced according to the same method used in experiment one.

5.1.2. Procedure. The procedure for experiment two was the same as that for experiment one.

5.1.3. Subjects. Five of the eight subjects who participated in experiment one also participated in experiment two (S1, S2, S4, S6, and S7). This was done in order to, in essence, hold a subset of the listeners' auditory systems constant so that direct comparisons across experiments (i.e., across syllable positions) could be made. The average age of participants was 21.5 (19-23). All were native speakers of English with, on average, 4.3 (3-6.5) years of second language study. All but one were right handed, and all but one were from the Midwest (the other was from the South). All participants were screened to ensure normal hearing.

5.1.4. Analysis. Data analysis was carried out as in experiment one (and as described in Section 3). Posterior distributions of Δ and π from a hierarchical multinomial-Dirichlet model were employed to evaluate marginal response invariance and report independence in identification-confusion data, and these results were used to infer patterns of perceptual and decisional independence or interaction.

5.2. Results

5.2.1. Model Checking. As noted above, the hierarchical multinomial-Dirichlet model has θ values equal in number (and corresponding to) each subject's identification-confusion frequencies, and so should fit the data quite well, although the group level in the hierarchical structure of the model constrains the individual level somewhat, so the fit need not be perfect. Figure 5.1 shows posterior distributions of θ values for S1. As is clear in the figure, the model fits extremely well (i.e., modeled identification-confusion probabilities – θ values – correspond closely to observed identification-confusion proportions). Thus, we may confidently employ the analyses of marginal response invariance and report independence presented below.

5.2.2. Analysis of marginal response invariance. Figure 5.2 shows the results of the analysis of marginal response invariance for the place of articulation dimension (i.e., Δ_{labial} and $\Delta_{alveolar}$). For half of the subjects – S1, S3, S4, and S6 – marginal response invariance fails for 'labial' responses (zero is outside the 99% highest probability density region). For three of these subjects – S1, S3, and S4 – marginal response invariance fails due to higher probability of 'labial' responses when presented with [p] than when presented with [b] (i.e., the posterior distributions of Δ_{labial} are shifted toward positive values), whereas for S6, the failure is due to the opposite tendency. Only

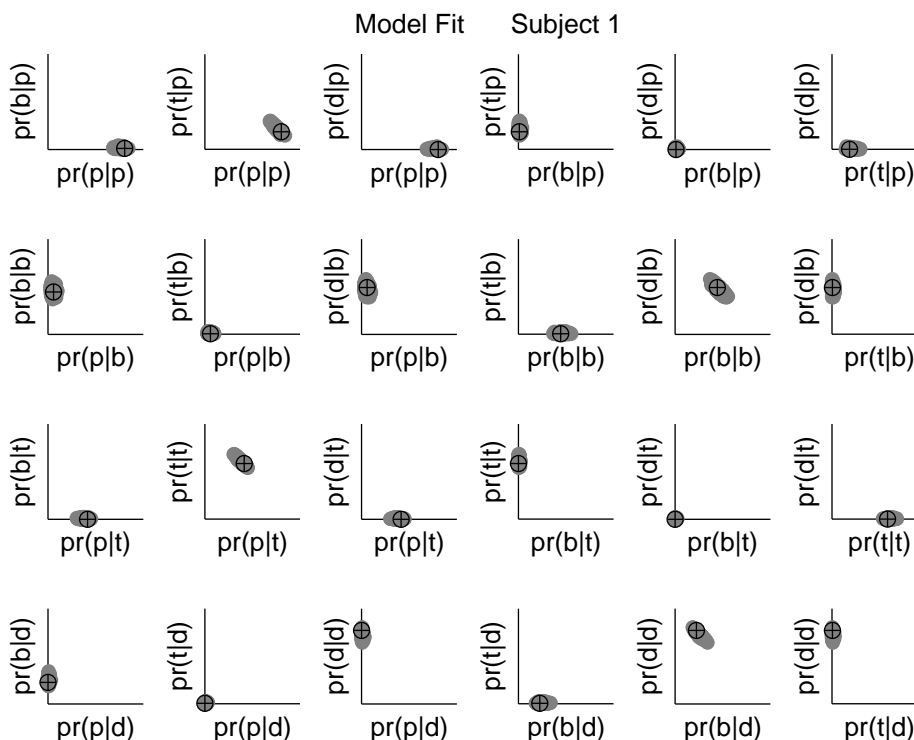


FIGURE 5.1. Scatter plots showing pairs of modeled identification-confusion probabilities (i.e., θ values; gray clouds) and observed identification-confusion proportions (black crosses) for subject one.

three subjects – S5, S7, and S8 – have posterior distributions of Δ_{labial} that have roughly as many positive as negative values (i.e., that have zero in a relatively high probability density region).

Only S1 exhibits a failure of marginal response invariance for ‘alveolar’ responses. The posterior distributions and observed values of $\Delta_{alveolar}$ for four subjects – S1, S3, S4, and S8 – show a large shift toward negative values, indicating higher probability of responding ‘alveolar’ when presented with [d] than when presented with [t]. Three subjects – S2, S6, and S7 – show the opposite tendency, and only S5’s posterior distribution has roughly equal proportions of $\Delta_{alveolar}$ on either side of zero. The group level analysis suggests that marginal response invariance holds for both ‘labial’ and ‘alveolar’ responses, though the distribution of Δ_{labial} consists mostly of positive values,

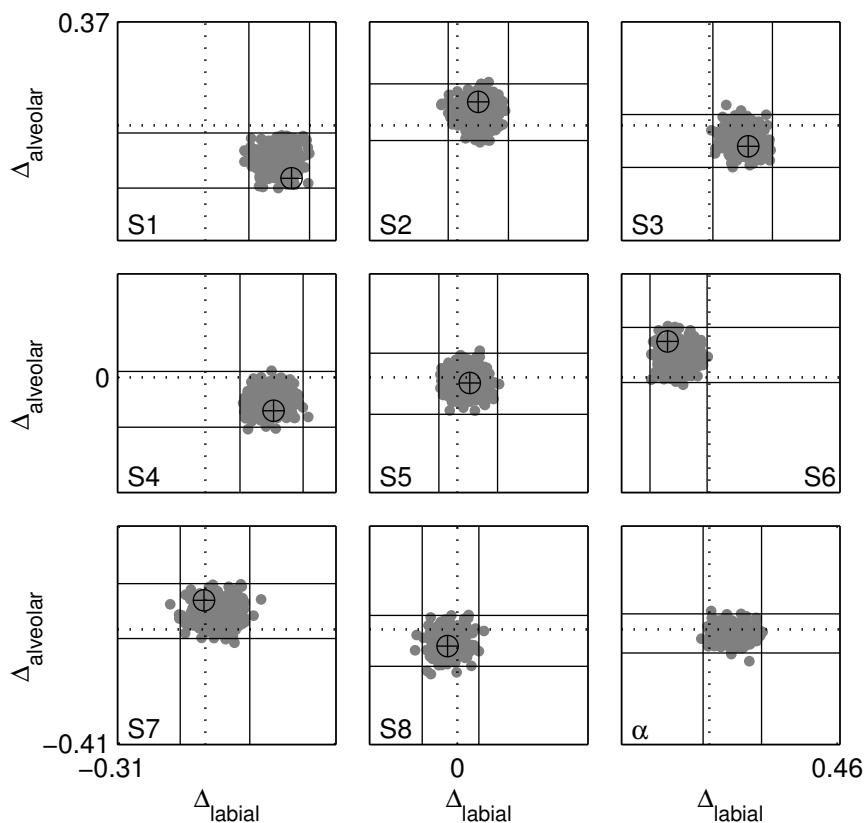


FIGURE 5.2. Scatter plots showing Δ_{labial} and $\Delta_{alveolar}$ (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of sampled posterior Δ values; observed statistics are indicated by black circled crosses. The x -axes indicate values of Δ_{labial} ; the y -axes indicate values of $\Delta_{alveolar}$. The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

again suggesting a greater tendency toward ‘labial’ responses in response to [p] than in response to [b].

Because marginal response invariance holds for at least one level on the place dimension (for all subjects but S1), these results are consistent with an underlying model in which decisional separability holds for the decision bound delineating ‘labial’ and ‘alveolar’ responses. For S3, S4, and S6, the failures of marginal response invariance for ‘labial’ responses suggests a change in salience

across levels of voicing. For S3 and S4, the change is a *decreased* salience of the place distinction for voiced (relative to voiceless) stops, as indicated by the positive shift in Δ_{labial} values, whereas for S6, the change is an *increased* salience of the place distinction for voiced (relative to voiceless) stops, as indicated by the negative shift in Δ_{labial} values. For S1, the shift in Δ_{labial} is positive and the shift in $\Delta_{alveolar}$ is negative, a pattern that can be accounted for by either mean-shift integrality wherein the [b] and [d] perceptual distributions are shifted toward the alveolar end of perceptual space relative to the [p] and [t] distributions, or by failure of decisional separability wherein the response regions for [p] and [d] are enlarged relative to [t] and [b], respectively.

Figure 5.3 shows the results of the analysis of marginal response invariance for ‘voiceless’ and ‘voiced’ responses. Only S4 exhibits a failure of marginal response invariance for ‘voiceless’ responses (i.e., a 99% highest probability density region that excludes zero). The posterior distributions of $\Delta_{voiceless}$ for five subjects – S1, S4, S5, S6, and S8 – are shifted, to varying degrees, toward negative values, indicating a higher probability of a ‘voiceless’ responses when presented with [t] than when presented with [p]. The other three subjects’ posterior distributions are fairly equally balanced on either side of zero.

Only two subjects – S1 and S2 – exhibit failure of marginal response invariance for ‘voiced’ responses (i.e., have 99% highest probability density regions that exclude zero), both with a shift toward negative values, indicating a higher probability of ‘voiced’ responses when presented with [d] than when presented with [b]. Only one subject’s Δ_{voiced} posterior distribution is shifted toward positive values (S3), and only two subjects’ posterior distributions have roughly equal proportions of positive and negative values (S6 and S8). The other five subjects all show substantial shifts toward negative values of Δ_{voiced} .

There are considerable differences in the variances of the posterior distributions of both $\Delta_{voiceless}$ and Δ_{voiced} across subjects. Three subjects – S3, S6, and S8 – have very low variability in the posterior distributions for both responses. At the other extreme, three subjects – S2, S4, and S5 – show much larger variability in these posterior distributions. For all subjects, the variances of the posterior distributions of Δ on each level of the voicing dimension are considerably smaller than variances of the posterior distributions of Δ on each level of the place dimension (the ranges of the axes in Figure 5.2 are more than twice as large as the ranges of the axes in Figure 5.3).

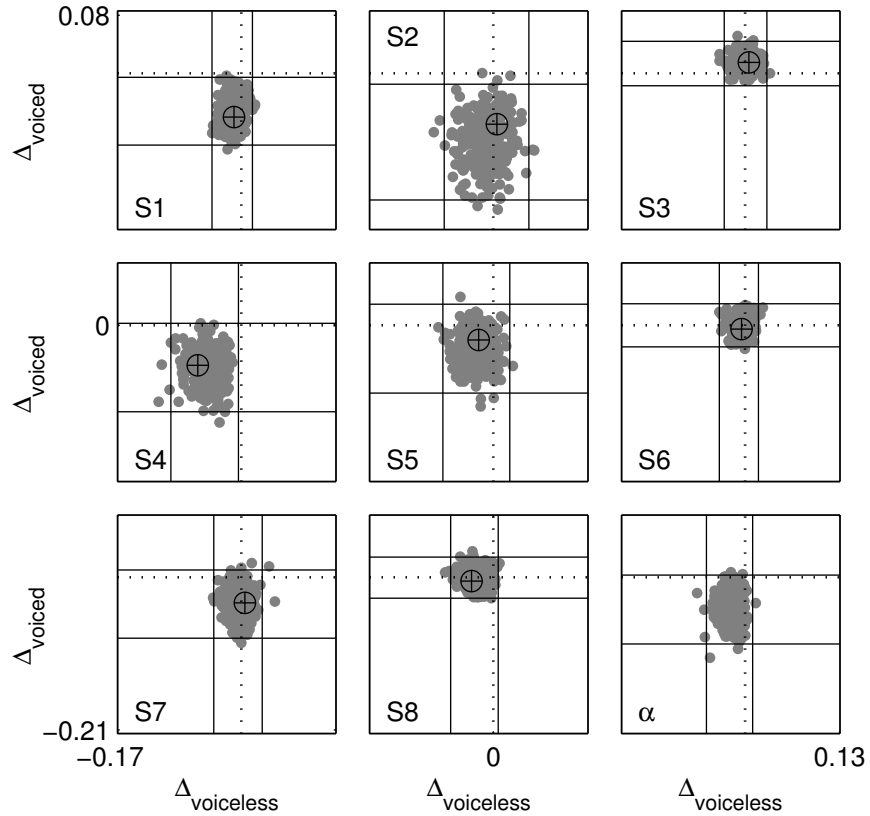


FIGURE 5.3. Scatter plots showing $\Delta_{voiceless}$ and Δ_{voiced} (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of sampled posterior Δ values (i.e., transformed posterior θ values); observed statistics are indicated by black circled crosses. The x -axes indicate values of $\Delta_{voiceless}$ (i.e., the probability of responding ‘voiceless’ when presented with [p] minus the probability of responding ‘voiceless’ when presented with [t]); the y -axes indicate values of Δ_{voiced} (i.e., the probability of responding ‘voiced’ when presented with [b] minus the probability of responding ‘voiced’ when presented with [d]). The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

The group level analysis suggests that marginal response invariance holds for both ‘voiceless’ and ‘voiced’ responses, though in each case, zero is very close to the limit of the 99% highest probability

density regions. There is a substantial shift toward negative values for both ‘voiced’ and ‘voiceless’ responses at the group level.

This pattern of results is consistent, for every subject, with the presence of decisional separability for the decision bound delineating ‘voiceless’ and ‘voiced’ responses. Given this, it is reasonable to infer that the three failures of marginal response invariance – for ‘voiceless’ responses for S4, and for ‘voiced’ responses for S1 and S2 – are due to failures of perceptual separability. For S1 and S2, this interaction appears to be due to higher salience of [d] than [b] (from [t] and [p], respectively), whereas for S4, this interaction appears to be due to higher salience of [t] than [p] (from [d] and [b], respectively). In all three cases, it seems to be the case that the voicing contrast was more salient for alveolar stops than for labial stops.

5.2.3. Analysis of report independence. Figure 5.4 shows the observed values and posterior distributions of π_p and π_b . Report independence holds for [p] for seven subjects and fails for one (S4). For this subject, given that the results of the analysis of marginal response invariance are consistent with the presence of decisional separability, the failure of report independence suggests that perceptual independence has failed in such a way that a larger portion of the [p] perceptual distribution falls in the correct response region than in the neighboring partially correct response regions.

Report independence fails for [b] for three subjects (S1, S2, and S4). For S1, the results of the analysis of marginal response invariance are consistent with a failure of decisional separability such that place of articulation responses depend on the level of voicing, so this subject’s failure of report independence may be due to either a failure of perceptual independence or decisional separability (or both). In either case, S1’s posterior distribution of π_b indicates (by its shift to negative values) that a smaller proportion of the [b] perceptual distribution falls in the correct response region than in the neighboring partially correct regions (relative to the case with both perceptual independence and decisional separability). S4’s failure of report independence is also due to negative π_b values. By way of contrast, the posterior distribution of π_b for S2 is shifted to positive values, which, given that marginal response invariance holds at all levels for S2, strongly suggests that perceptual independence has failed by shifting relatively more of the [b] perceptual distribution to fall in the correct than in the partially correct response regions.

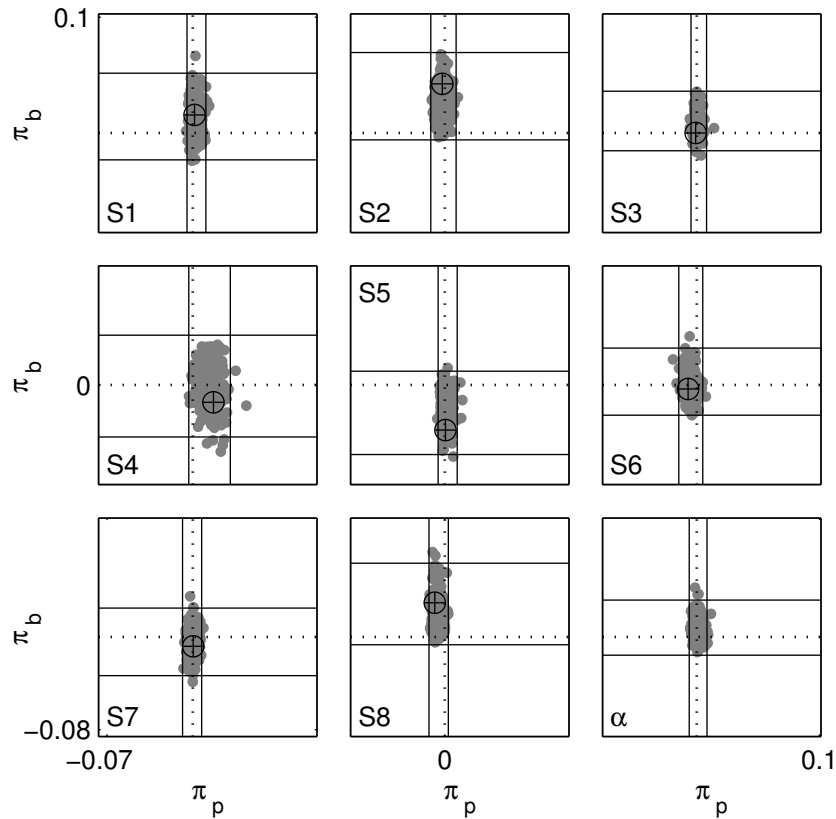


FIGURE 5.4. Scatter plots showing π_p and π_b (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for labial stimuli [p] and [b]. The x -axes indicate the value of π_p ; the y -axes indicate the value of π_b . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

For three subjects – S2, S4, and S5 – the posterior distributions of π_p and π_b exhibit considerably larger variance than do those of the other five subjects. The group level analysis shows fairly low variance for both π_p and π_b and indicates that report independence holds for both stimuli at the group level.

Figure 5.5 shows the results of the analysis of report independence for [t] and [d]. Report independence holds for [t] for all but one subject (S2), who shows a positive shift of π_t values.

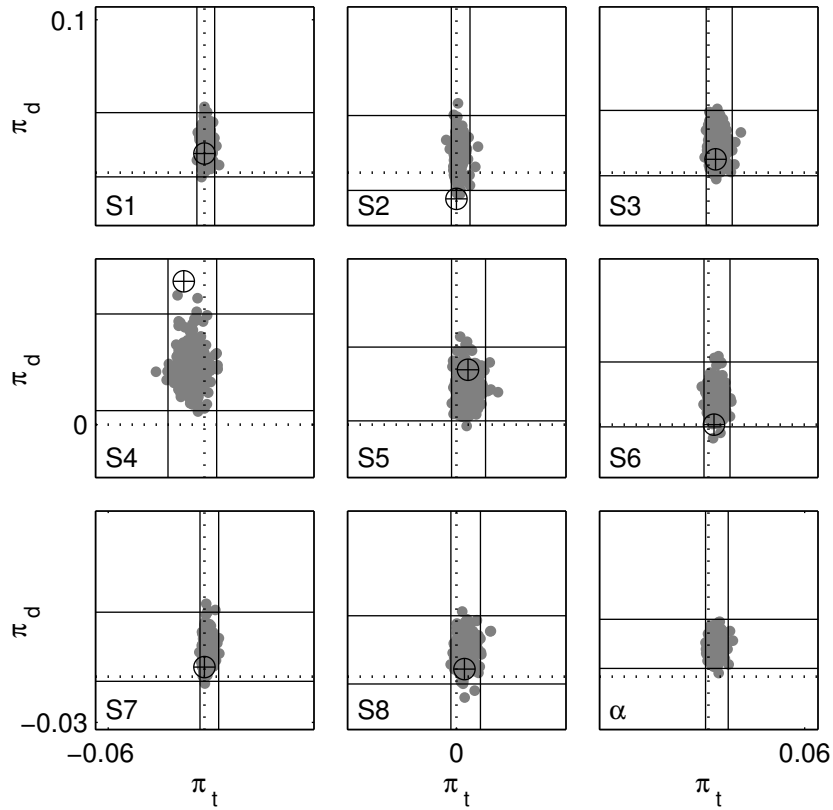


FIGURE 5.5. Scatter plots showing π_t and π_d (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for labial stimuli [t] and [d]. The x -axes indicate the value of π_t ; the y -axes indicate the value of π_d . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

Again, marginal response invariance holds for S2, suggesting that decisional separability holds for each dimension, which in turn suggests that this failure of report independence is due to a failure of perceptual independence. The positive shift indicates that relatively more of the perceptual distribution is located in the correct response region than in the neighboring partially correct regions.

Report independence holds for [d] for six subjects and fails for two. For both S4 and S5, the presence of marginal response invariance suggests that decisional separability holds, suggesting that

the failures of report independence are due to failures of perceptual independence. In each case, the observed and modeled values of π_d are shifted toward positive values, indicating relatively more of the perceptual distribution in the correct response region than in the adjacent partially correct regions.

The group level posterior distributions of π_t and π_d indicate that report independence holds for [t] and fails for [d]. In each case, there is a strong positive shift in the distributions of π .

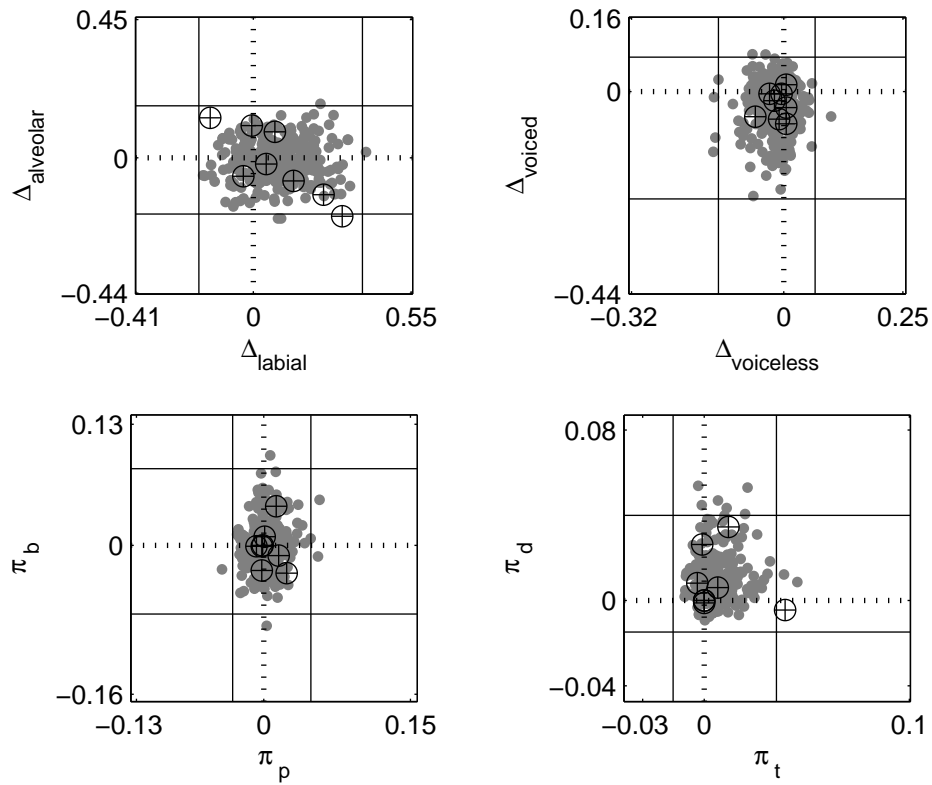


FIGURE 5.6. Scatter plots showing predicted and observed Δ and π statistics. The gray clouds show a subset (1/10) of the predicted values of Δ and π generated from the posterior distribution of the group level α ‘counts’. Black circled crosses indicate observed values of Δ and π . Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

5.2.4. Prediction of future data. Figure 5.6 shows predicted and observed (for all eight subjects) values of Δ statistics for each level on each dimension and π statistics for each stimulus. Given

the observed identification-confusion data, the added uncertainty (due to simulation of data based on group level modeled α values), the demographic properties of the subjects, and the experimental materials and protocol, the Bayesian model predicts that marginal response invariance will hold for both levels on each dimension – ‘labial’, ‘alveolar’, ‘voiced’, and ‘voiceless’ responses – and that report independence will hold for all four consonants.

The predicted distributions of Δ_{labial} and $\Delta_{alveolar}$ are roughly evenly balanced about zero, whereas the predicted distributions of Δ_{voiced} and $\Delta_{voiceless}$ both exhibit shifts toward negative values. These negative shifts indicate that correct voicing responses are expected to be more likely in response to alveolar [t] and [d] than to labial [p] and [b].

The predicted distributions of π_p and π_b are roughly evenly distributed around zero, while the predicted distributions of π_t and π_d exhibit substantial positive shifts. These shifts indicate that relatively more of the [t] and [d] distributions are expected to fall in the correct response regions than in the neighboring partially correct regions.

TABLE 5.1. Summary of the results of analyses of marginal response invariance (Δ) and report independence (π). $p(C)$ = proportion correct responses. *lab.* = labial ([p] or [b]); *alv.* = alveolar ([t] or [d]); *-voi* = voiceless ([p] or [t]); *+voi* = voiced ([b] or [d]). \checkmark indicates that marginal response invariance or report independence holds (i.e., that zero is within the 99% highest probability density region) for a given subject for that level and dimension (Δ) or stimulus (π).

Sub.	$p(C)$	$\Delta_{lab.}$	$\Delta_{alv.}$	Δ_{-voi}	Δ_{+voi}	π_p	π_b	π_t	π_d
1	0.65			\checkmark		\checkmark		\checkmark	\checkmark
2	0.57	\checkmark	\checkmark	\checkmark		\checkmark			\checkmark
3	0.71		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
4	0.54		\checkmark		\checkmark			\checkmark	
5	0.61	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	
6	0.71		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
7	0.58	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
8	0.74	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
α		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	

5.2.5. Summary of results. Table 5.1 presents a summary of the discrete inferences drawn from the analyses of marginal response invariance and report independence (i.e., inferences based on 99% highest probability density region limits), as well as the proportion of correct responses for each subject. Accuracy is somewhat lower for coda position than it is for onset position, ranging from 54% to 74% correct. The results of the analyses of marginal response invariance suggest that for all but one subject – S1 – decisional separability holds. For three subjects – S3, S4, and S6 – observed failures of marginal response invariance suggest underlying failures of perceptual separability. For S3 and S4, these failures appear to be due to decreased salience of the place distinction for voiced [b] relative to voiceless [p] (from [d] and [t], respectively); for S6 the pattern is reversed, such that voiceless [p] is more salient than is [b] (from [t] and [d], respectively). S1 exhibits a pattern of failures of marginal response invariance consistent with failure of decisional separability or (mean-shift-type) failure of perceptual separability (or both). The group level analyses indicate that marginal response invariance holds for each level on each dimension. Not surprisingly, then, the model predicts the same.

Given the apparent presence of decisional separability (for all subjects other than S1), the results of the Bayesian analyses of report independence suggest that the small number of failures of report independence are due to failures of perceptual independence (i.e., perceptual correlations or cross talk between place and voicing processing channels). The group level analysis indicates that report independence holds for all four stimuli.

Thus, aside from S1, the results for syllable coda stop consonants indicate that place of articulation and voicing tend fairly strongly to generalize across one another in both perception and decision making and that place and voicing channels are independent.

5.3. Discussion

To the extent that the explanations for failures of perceptual separability (inferred via failure of marginal response invariance) in experiment one are correct, it should not be surprising that the same consonants in syllable coda position do not show the same kinds or degree of perceptual interaction. The juncture between consonant and vowel is not delimited by a burst in coda stops the way it is in onset stops, so differences in burst strength between labials and alveolars cannot induce a failure of perceptual separability. Similarly, there is not formant cutback in coda stops, so

there is, in principle, no difference in the amount of place of articulation information encoded in the formant transitions leading up to coda stop closures. In any case, any discussion of the cause(s) of the observed patterns is speculative; determining the cause(s) of the observed patterns is beyond the scope of this work.

Determining that the differences in presence or absence of perceptual interactions between onset and coda stops is robust, on the other hand, is well within the scope of this work. Recall that five of the eight subjects from this experiment also participated in experiment one (S1, S2, S4, S6 and S7). Comparing these subjects' results across the two experiments strengthens our confidence that any differences between the two experiments are due to changes in the acoustic-phonetic structure of the consonants due to changes in syllable position rather than linguistically irrelevant or uninteresting differences between two unrelated groups of experimental subjects.

Table 5.2 shows a comparison of these five subjects' marginal response invariance and report independence results for onset and coda [p], [b], [t], and [d]. Plus signs indicate the same test result for each experiment; minus signs indicate different test results across experiments. For example, for S1, marginal response invariance holds for 'labial' responses for onset stimuli and fails for coda stimuli; fails for 'alveolar' responses for both onset and coda stimuli; and holds for 'voiceless' responses for both onset and coda stimuli. Thus, the first three columns for S1 in the table are $-$, $+$, and $+$.

TABLE 5.2. Summary of the differences between onset and coda stop interactions.

$+/-$ indicate the same or different results of tests of marginal response invariance (Δ) and report independence (π) across experiments, respectively.

Sub.	$\Delta_{lab.}$	$\Delta_{alv.}$	Δ_{-voi}	Δ_{+voi}	π_p	π_b	π_t	π_d
1	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$
2	$-$	$-$	$-$	$-$	$+$	$+$	$-$	$+$
4	$-$	$-$	$-$	$-$	$+$	$-$	$-$	$-$
6	$+$	$-$	$+$	$-$	$+$	$+$	$+$	$-$
7	$-$	$-$	$+$	$+$	$+$	$+$	$+$	$+$
α	$+$	$-$	$+$	$-$	$+$	$+$	$+$	$+$

Two characteristics of the comparison between onset and coda stimuli are immediately apparent. First, 14 out of 20 tests of marginal response invariance are different for onset vs. coda stimuli. For

two subjects – S2 and S4 – none of the marginal response invariance results are the same; for the other three, half of each subjects' results differ across syllable positions. Second, 13 of 20 tests of report independence are the same for onset and coda stimuli. One subject – S4 – exhibits three (of four possible) differences, and one – S1 – exhibits two differences, but the other three share most or all of the report independence test results for both syllable positions. The comparison of group level analyses across syllable positions shows a similar pattern, namely a tendency toward disagreement with respect to marginal response invariance and toward agreement with respect to report independence.

It seems to be the case that place and voicing tend strongly to generalize across one another with regard to decision making for both onset and coda stimuli (i.e., decisional separability seems to hold for most subjects on both dimensions for both syllable positions). On the other hand, place and voicing tend to generalize perceptually across one another much more for coda stops than for onset stops (i.e., perceptual separability seems to hold for coda stops and fail for onset stops). Finally, there is relatively little interaction between processing channels within stimuli for either onset or coda stimuli (i.e., perceptual independence tends to hold for all four stimuli for both syllable positions).

Manner of articulation and voicing in onset position

Phonological space often makes use of more than two dimensions. For example, while voicing and place of articulation define a two dimensional subspace of the phonological space of English (among other languages), a partially overlapping two dimensional space can be defined in terms of voicing and manner of articulation. It is to this portion of phonological space that we now turn. More specifically, we will employ the experimental and analytic methods described above to two investigations of the integration of voicing and manner of articulation in the set of (English) obstruent consonants [p], [b], [f], and [v] in onset and coda position. Figure 6.1 illustrates the abstract relationship between the voicing and place of articulation space studied above and the voicing and manner of articulation space that is the focus of this and the following chapter. In this chapter, we consider the perceptual and decisional relationships between voicing and manner in onset position, and in the following chapter we do so in coda position.

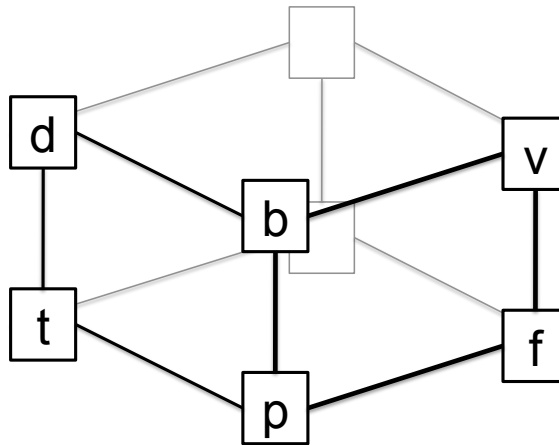


FIGURE 6.1. Schematic representation of the structure of voicing, place of articulation, and manner of articulation in a subset of (English) phonological space.

As discussed above, the distinction between voiced and voiceless stops is cued by differences on multiple acoustic-phonetic dimensions, among them differences in VOT, formant cutback, and

release burst amplitude. The distinction between voiced and voiceless fricatives consists of a different set of multiple acoustic-phonetic cues. For example, voiceless fricatives tend to be longer in duration, have less acoustic power at low frequencies (i.e., those excited by regular oscillation of vocal folds), and have more acoustic power at high frequencies than do their voiced counterparts (e.g., Jongman, Wayland, & Wong, 2000; Silbert & Jong, 2008).

The distinctions between stops and fricatives are similarly multidimensional. A voiceless (English) stop consists of a release burst followed by low amplitude aspiration noise in the place of formant transition information (i.e., cutback), whereas a voiceless fricative consists of continuous noise with largely consistent levels and spectral properties (relative to the variation in an analogous stop), and no cutback of vowel formants. On the other hand, voiced stops consist of relatively low level release bursts, some degree of voicing preceding release or only a short VOT after, whereas a voiced fricative consists of continuous noise, usually with some degree of simultaneous voicing, typically near the consonant-vowel boundary.

The point here is not to provide even an approximately exhaustive catalog of the acoustic-phonetic cues relevant to voicing and manner of articulation in labial stops and fricatives. Rather, it is to provide a brief illustration of the complex mapping between multiple acoustic-phonetic dimensions and multiple phonological dimensions. It is difficult to state with any certainty what kinds of patterns of perceptual and decisional independence or interaction to expect, though we may surmise that differences in voicing and place cues for stops and fricatives will induce perceptual interactions.

6.1. Methods

6.1.1. Stimuli. Four tokens of each stimulus type – [pa], [ba], [fa], [va] – were produced by the first author (the [pa] and [ba] tokens were the same tokens employed in experiment one). The stimuli for both experiments were recorded on the same equipment and during the same session.

6.1.2. Procedure. The procedure was identical to that employed in experiment one, with appropriate changes made to instructions and button labels.

6.1.3. Subjects. Eight adults (three male, five female) were recruited from the university community. The average age of participants was 22 (18-27). All were native speakers of English with, on average, 5.75 (2-7) years of second language study. All but two were right handed, and all

but two were from the Midwest (the other two were from the East). All participants were screened to ensure normal hearing.

6.1.4. Analysis. Analyses were carried out in the same manner as those in the previous experiments.

6.2. Results

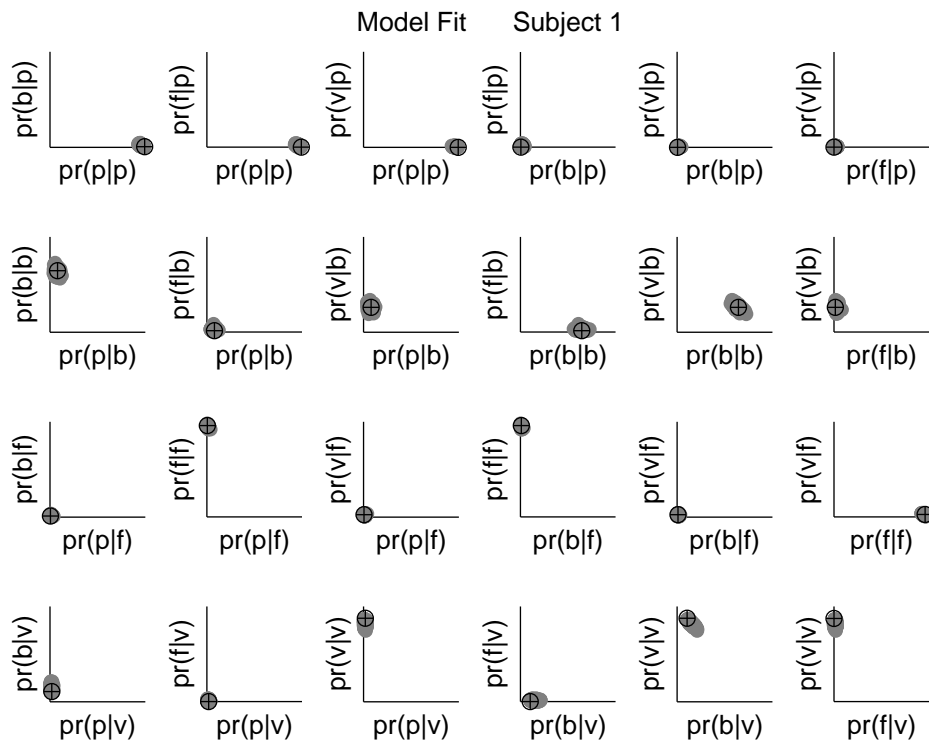


FIGURE 6.2. Scatter plots showing pairs of modeled identification-confusion probabilities (i.e., θ values; gray clouds) and observed identification-confusion proportions (black crosses) for subject one.

6.2.1. Model checking. Figure 6.2 shows observed identification-confusion proportions and posterior distributions of θ values for each pair of responses for each stimulus for subject one. The model fits every subjects data similarly well, so we can safely make use of posterior distributions

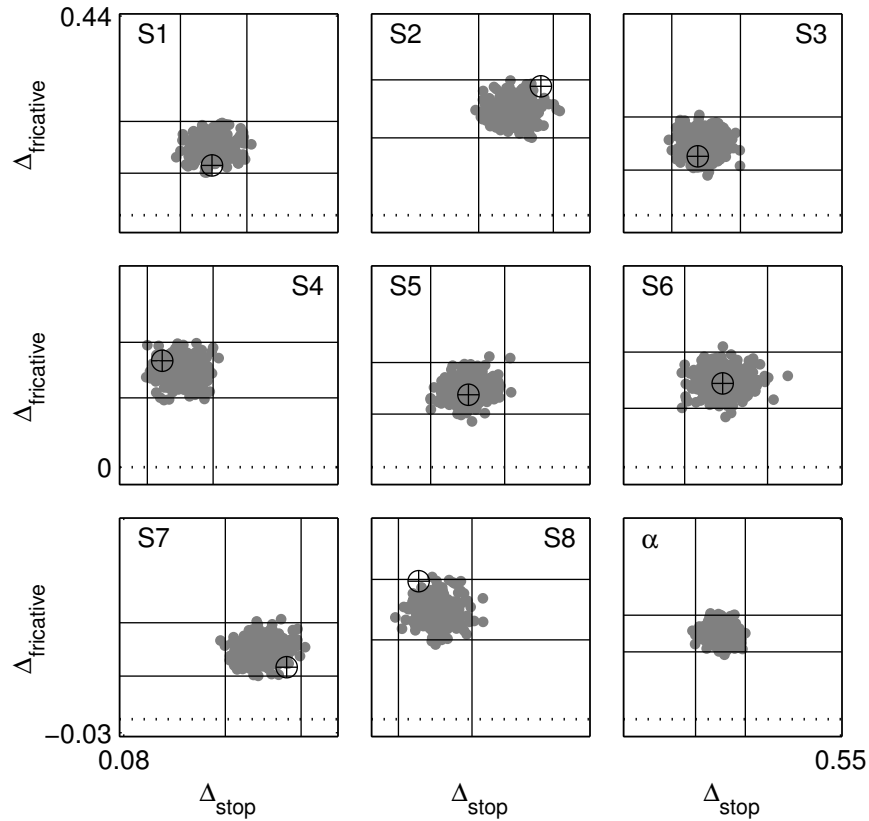


FIGURE 6.3. Scatter plots showing Δ_{stop} and $\Delta_{fricative}$ (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of sampled posterior Δ values; observed statistics are indicated by black circled crosses. The x -axes indicate values of Δ_{stop} ; the y -axes indicate values of $\Delta_{fricative}$. The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

of Δ and π statistics to draw inferences about patterns of perceptual and decisional interactions or independence.

6.2.2. Analysis of marginal response invariance. Figure 6.3 shows a subset of the samples from the posterior distributions of Δ_{stop} (x -axis) and $\Delta_{fricative}$ (y -axis). For all subjects, without exception, the posterior distributions of both statistics are shifted to positive values. With regard to Δ_{stop} , this indicates that subjects were more likely to respond ‘stop’ (‘p’ or ‘b’) when presented

with [p] than when presented with [b], and with regard to $\Delta_{fricative}$, this indicates that subjects were more likely to respond ‘fricative’ (‘f’ or ‘v’) when presented with [f] than when presented with [v]. The positive shifts in both statistics is consistent with a model in which decisional separability holds and perceptual separability fails due to a change in salience of manner of articulation across levels of voicing; the manner distinction between [p] and [f] is more salient than it is between [b] and [v].

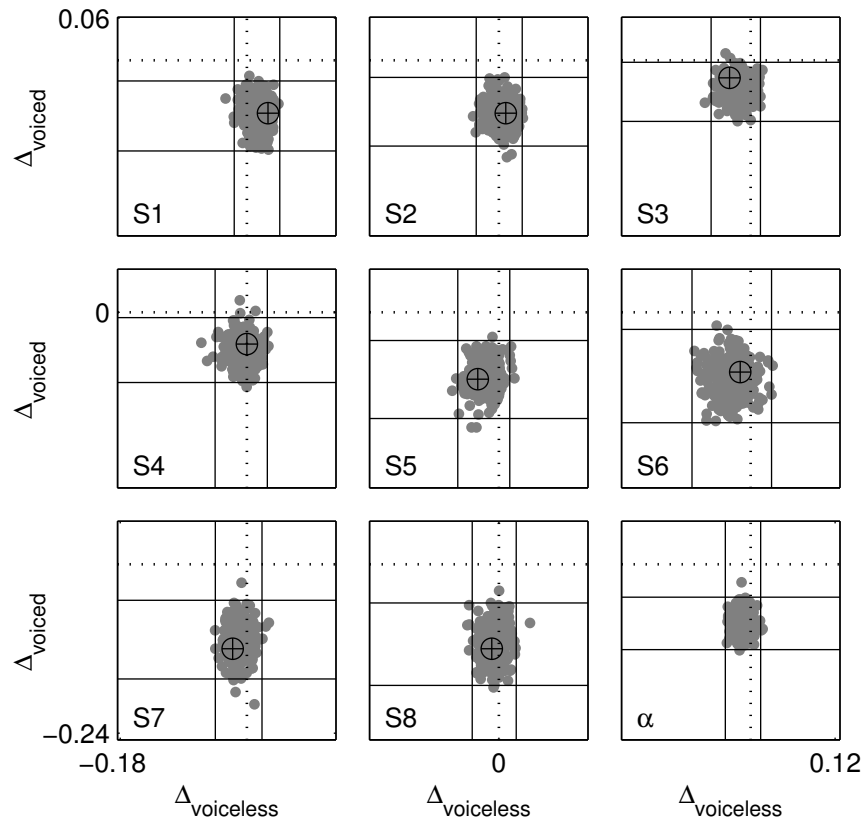


FIGURE 6.4. Scatter plots showing $\Delta_{voiceless}$ and Δ_{voiced} (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of sampled posterior Δ values; observed statistics are indicated by black circled crosses. The x -axes indicate values of $\Delta_{voiceless}$; the y -axes indicate values of Δ_{voiced} . The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

All eight subjects exhibit similar variances in the posterior distributions of Δ_{stop} and $\Delta_{fricative}$, though there is some variability in the magnitude of the positive shift of these statistics between subjects. For example, S2 exhibits a large positive shift for both Δ_{stop} and $\Delta_{fricative}$, while S4 and S8 have much larger positive shifts for $\Delta_{fricative}$ than for Δ_{stop} . The rest of the subjects exhibit similar, more moderate positive shifts for each statistic. The group level analysis shows the same pattern, though with fairly little variability on either Δ_{stop} or $\Delta_{fricative}$.

Figure 6.4 displays the posterior distributions of $\Delta_{voiceless}$ and Δ_{voiced} . There is again complete consistency across subjects. Zero is well within the 99% highest probability density region for the posterior distribution of $\Delta_{voiceless}$, indicating that marginal response invariance holds, for all eight subjects. For one subject – S1 – the observed value of $\Delta_{voiceless}$ is positive, and the posterior distribution of $\Delta_{voiceless}$ is shifted toward positive values. For four subjects – S3, S5, S6, and S7 – the opposite is true; for each, observed $\Delta_{voiceless}$ is negative and the posterior distribution of $\Delta_{voiceless}$ consists of mostly negative values. For the other three subjects, the observed values of $\Delta_{voiceless}$ are very near zero and the posterior distributions have roughly equal proportions of positive and negative values. The negative shift indicates an increased probability of responding ‘voiceless’ when presented with [f] than when presented with [p]; the positive shift indicates the opposite. On the other hand, the posterior distributions of Δ_{voiced} are shifted to negative values for every subject, indicating that subjects were more likely to respond ‘voiced’ (‘b’ or ‘v’) when presented with [v] than when presented with [b]. Zero is excluded from the 99% highest probability density region for all eight subjects, strongly suggesting that marginal response invariance fails across the board.

Taken together, the presence of marginal response invariance for ‘voiceless’ responses suggests that voicing is decisionally separable from manner, so the failure of marginal response invariance for ‘voiced’ responses indicates that perceptual separability fails because [v] is perceptually more ‘voiced’ than is [b].

The group level analysis shows the same pattern, indicating that marginal response invariance holds for ‘voiceless’ responses (with a tendency toward negative values) and fails for ‘voiced’ responses. In both the individual and group level analyses, the variances of the posterior distributions are smaller for the voicing dimension than the manner dimension.

6.2.3. Analysis of report independence. Figure 6.5 shows the posterior distributions of π_p (x -axes) and π_b (y -axes). For all subjects other than S7, zero is within the highest probability density region for π_p . In every case, the posterior distributions consist almost entirely of positive values, indicating that relatively more of the perceptual distribution for [p] falls in the correct response region than in the neighboring partially correct regions. This positive shift may be due to the high accuracy rates in this experiment (see Table 6.1, below), or it may be due to ‘near-miss’ failures of perceptual independence in subjects’ [p] perceptual distributions.

For all but two subjects – S7 and S8 – zero is within the highest probability density region for π_b . For all subjects other than S1 (and perhaps S3), the bulk of the posterior distributions consist of positive values, again indicating relatively more of the [b] perceptual distributions falling in the correct than in the partially correct response regions.

In the group level analysis, zero falls just outside the 99% probability density region for π_p , and just inside the analogous region for π_b . In both cases, as with the individual subject analyses, the bulk of the posterior samples are positive.

Figure 6.6 shows the posterior distributions of π_f (x -axes) and π_v (y -axes). Zero falls outside the highest probability density region of π_f , indicating that report independence fails for [f], for all subjects other than S2. For all eight subjects, the posterior distributions are shifted toward positive values. Given that the analyses of marginal response invariance strongly suggest that decisional separability holds for both dimensions, the failures of report independence here suggests that perceptual independence fails for the [f] perceptual distribution such that, relative to a case in which perceptual independence holds, more of the [f] perceptual distribution falls in the correct than in the adjacent partially correct response regions.

On the other hand, report independence holds, suggesting that perceptual independence holds, for all subjects (other than S1) for [b]. For two of the subjects – S2 and S4 – the posterior distributions of π_b have roughly equal proportions of positive and negative values, whereas for the other six subjects, the posterior distributions are shifted toward positive values.

The group level analysis is consistent with the individual subject level analyses. Report independence fails for [f] and holds for [v], and both π_f and π_v are shifted towards positive values.

6.2.4. Prediction of future data. Figure 6.7 shows the predicted and observed values of Δ for each level of each dimension and π statistics for each stimulus for this experiment. The predicted

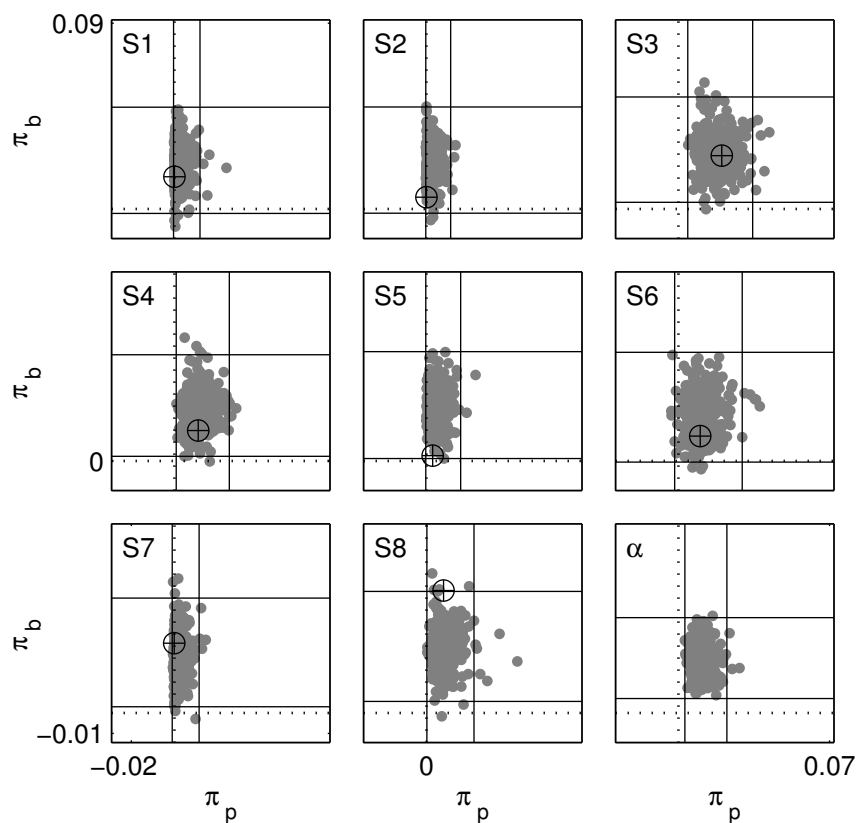


FIGURE 6.5. Scatter plots showing π_p and π_b (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for stop stimuli [p] and [b]. The x -axes indicate the value of π_p ; the y -axes indicate the value of π_b . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

distributions of Δ statistics indicate that we should expect marginal response invariance to fail at both levels of manner of articulation due to higher accuracy (presumably due to greater salience) for [p] and [f] relative to [b] and [v]. Marginal response invariance is predicted to hold at both levels of voicing, the the predicted distribution of Δ_{voiced} consists almost entirely of negative values, indicating a greater predicted probability of responding ‘voiced’ when presented with [v] than when

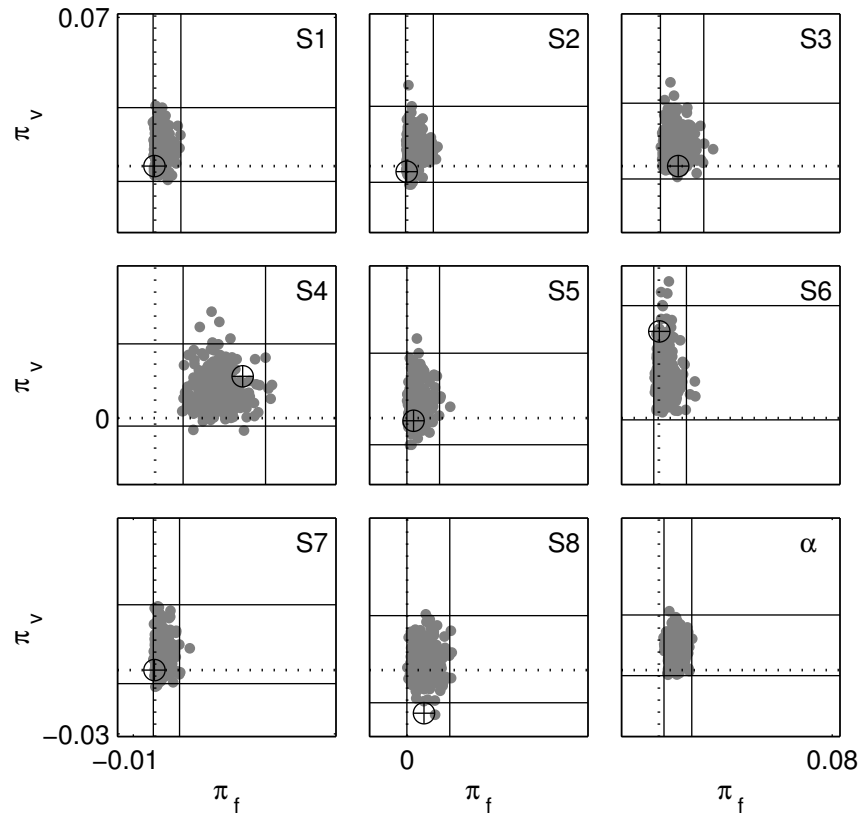


FIGURE 6.6. Scatter plots showing π_f and π_v (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for fricative stimuli [f] and [v]. The x -axes indicate the value of π_f ; the y -axes indicate the value of π_v . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

presented with [b]. The spread of observed and predicted values of Δ_{stop} and $\Delta_{fricative}$ are much greater than are the observed and predicted values of $\Delta_{voiceless}$ and Δ_{voiced} .

The predicted distributions of π indicate that we should expect report independence to fail for [f] and hold for [p], [b], and [v], with largely positive values of π for [f] and [p] and a somewhat more balanced distributions of positive and negative values of π for [v] and [b].

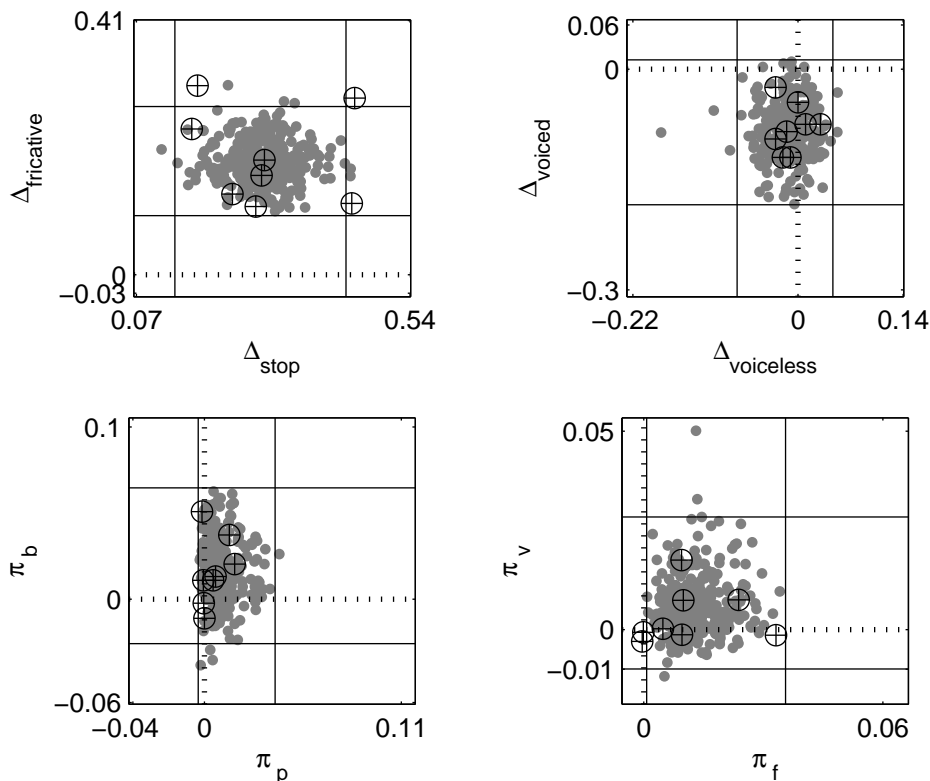


FIGURE 6.7. Scatter plots showing predicted and observed Δ and π statistics. The gray clouds show a subset (1/10) of the predicted values of Δ and π generated from the posterior distribution of the group level α ‘counts’. Black circled crosses indicate observed values of Δ and π . Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

6.2.5. Summary of results. Table 6.1 provides a summary of the discrete inferences drawn from the analyses presented above (i.e., inferences based on 99% highest probability density region limits) and accuracy rates for manner and voicing in onset position. Accuracy is very high here, ranging from 70% to 88% correct. All eight subjects and the group level analysis exhibit complete consistency with respect to marginal response invariance. Marginal response invariance holds for ‘voiceless’ responses and fails for ‘voiced’, ‘stop’, ‘fricative’ responses. In every case, the posterior distributions of Δ_{stop} and $\Delta_{\text{fricative}}$ consisted entirely of positive values, suggesting that manner is decisionally separable from voicing and indicating that perceptual separability fails due to greater

salience of [p] and [f] relative to [b] and [v]. The presence of marginal response invariance for ‘voiceless’ responses suggests that voicing is also decisionally separable from manner, and the mostly negative values in the posterior distributions of Δ_{voiced} indicate that [v] is more distinct from [f] than [b] is from [p].

There is only slightly less consistency across subjects in the analyses of report independence. For most subjects, report independence fails for [f] but holds for [p], [b], and [v]. Because the results of the analyses of marginal response invariance suggest that decisional separability holds across the board, the failures of report independence for [f] are likely due to failures of perceptual independence. In each case, the posterior distributions of π_f are shifted toward positive values, suggesting that relatively more of the [f] perceptual distributions fall in the correct response regions than in the neighboring partially correct regions. The group level analysis of report independence is consistent with the individual subject analyses for [b], [f], and [v], but exhibits a failure of report independence for [p]. Recall that, although zero was within the 99% highest probability density region of π_p for seven of eight subjects, the posterior distribution was shifted heavily toward positive values for every subject, suggesting that the evidence for report independence – and, thereby, perceptual independence – for [p] is somewhat tenuous.

6.3. Discussion

As with place of articulation and voicing, it seems that decisional separability holds mutually between manner of articulation and voicing in onset position. To the extent that GRT accurately captures the distinction between perception and decision making, it seems that (at least some) phonological features generalize across each other at post-perceptual levels of processing. In terms of perception, however, the implementation of phonological structure is not so straightforward.

The fact that manner of articulation is not perceptually separable from voicing is not especially surprising, at least in retrospect. What we might call purely consonantal acoustic-phonetic cues to the manner contrast (e.g., stop release burst and noise amplitude, noise duration) are ‘more distinct’ in the voiceless pair [p] and [f] than they are in the voiced pair [b] and [v]. Voiceless [p] has a relatively long VOT filled with aspiration noise, whereas voiced [b] has a short or even negative VOT (i.e., pre-voicing) and a relatively quiet release burst. Voiceless [f] has longer and

TABLE 6.1. Summary of the results of analyses of marginal response invariance (Δ) and report independence (π). $p(C)$ = proportion correct responses. *stop* = [p] or [b]; *fric.* = fricative ([f] or [v]); *-voi* = voiceless ([p] or [f]); *+voi* = voiced ([b] or [v]). \checkmark indicates that marginal response invariance or report independence holds (i.e., that zero is within the 99% highest probability density region) for a given subject for that level and dimension (Δ) or stimulus (π).

Sub.	$p(C)$	Δ_{stop}	$\Delta_{fric.}$	Δ_{-voi}	Δ_{+voi}	π_p	π_b	π_f	π_v
1	0.87			\checkmark		\checkmark	\checkmark		
2	0.78			\checkmark		\checkmark	\checkmark	\checkmark	\checkmark
3	0.88			\checkmark		\checkmark	\checkmark		\checkmark
4	0.85			\checkmark		\checkmark	\checkmark		\checkmark
5	0.80			\checkmark		\checkmark	\checkmark		\checkmark
6	0.70			\checkmark		\checkmark	\checkmark		\checkmark
7	0.80			\checkmark					\checkmark
8	0.81			\checkmark		\checkmark			\checkmark
α				\checkmark			\checkmark		\checkmark

louder frication noise than does [v]; vocal fold oscillation during frication reduces the amplitude and duration of the fricative noise (Stevens, Blumstein, Glicksman, Burton, & Kurowski, 1992).

The failure of voicing to be perceptually separable from manner, and the particular way in which this failure is manifested, is perhaps a bit more surprising. The presence of marginal response invariance for ‘voiceless’ responses is consistent with perceptual separability at this level, indicating that [p] and [f] percepts are equally, or at least very similarly, voiceless. The failure of marginal response invariance for ‘voiced’ responses indicates that [v] is perceptually ‘more voiced’ than [b].

The results of the analyses of report independence are not as transparently mapped onto to phonological structure. The apparent failure of perceptual independence for [f] stimuli and the consistent pattern of near-failures for [p] can be interpreted as processing dependencies between the voicing and manner channels when the stimuli are voiceless. On the other hand, the tendency for report independence to hold for both [b] and [v] suggests independence (or lower magnitude dependencies) between these channels when the stimuli are voiced. Failure of perceptual independence

has been discussed as ‘the strongest form of holism’ (Wenger & Ingvalson, 2003). Phonologically, we might tentatively think of a failure of perceptual independence as a fusing of features at particular levels, creating something akin to a perceptual segment.

Manner of articulation and voicing in coda position

Experiment four is an investigation of perceptual independence, perceptual separability, and decisional separability between manner of articulation and voicing in consonants in coda position. As in experiments one and two with regard to place and voicing, at an abstract level of analysis, the (phonological) structure of manner and voicing are, in some sense, independent of syllable structure. Stops are stops and fricatives are fricatives regardless of whether they precede or follow vowels (or other consonants); the same can be said of voiced and voiceless segments. Of course, the implementation of the phonological structure is likely to depend substantially on the phonological and phonetic context in which a consonant appears.

As noted above, the acoustic-phonetic structure of voicing (contrasts) exhibits fairly dramatic differences in onset and coda stops. VOT, release bursts, and (lack of) aspiration are among the relevant cues in onset stops, whereas vowel and consonant closure duration likely do the lion's share of work in coda position. It seems reasonable to posit that the cues to voicing contrasts in fricatives will not change as much with syllable position. Whereas differences in VOT and associated aspiration noise differentially impinge on vowel information in onset [p] and [b], the cues to voicing in [f] and [v] are more localized in the consonants themselves. Essentially the same cues (e.g., presence/absence of voicing, noise level and duration) may plausibly occur in [f] and [v] in coda position, whereas, as noted, the voicing cues in coda stops differ substantially. On the other hand, manner distinctions in coda position are likely largely cued by differences in the presence and amplitude of noise. The acoustic properties of fricatives are very similar in onset and coda position, whereas they differ considerably in stops.

It is difficult, if not impossible, to predict *a priori* how differences in the acoustic-phonetic structure will (or will not) induce changes in the pattern of perceptual and decisional interactions observed. Nonetheless, it is worth discussing these differences, if only in brief, in order to guide our thinking about possible causes for any such observed changes. We may hypothesize that, based on

the different pattern of results in experiments one and two, that coda position tends to induce better feature-based generalization (i.e., more instances of marginal response invariance holding). If so, we should expect a similar change here.

7.1. Methods

7.1.1. Stimuli. Four tokens of each stimulus type – [ap], [ab], [af], [av] – were produced by the first author (the [ap] and [ab] tokens were the same tokens employed in experiment two). As noted above, the stimuli for each experiment were recorded on the same equipment and during the same session.

7.1.2. Procedure. The procedure was identical to that employed in experiment one, with appropriate changes made to instructions and button labels.

7.1.3. Subjects. Eight adults (three male, five female) were recruited from the university community. The average age of participants was 21.8 (18-27). All were native speakers of English with, on average, 4.75 (2-9) years of second language study. All but two were right handed, and all but one were from the midwest (the other was from the east coast). All participants were screened to ensure normal hearing. Six of the subjects that participated in this experiment also participated in the previous experiment (i.e., manner and voicing in onset position).

7.1.4. Analysis. Analyses were carried out in the same manner as those in the previous experiments.

7.2. Results

7.2.1. Model checking. Figure 7.1 shows the observed and modeled identification-confusion probabilities from Subject 1 of the present experiment. As in the previous experiments, the model provides a statistically sound characterization of the data. Thus, we may proceed to draw inferences regarding perceptual and decisional interactions based on the analysis of marginal response invariance and report independence with a reasonably high degree of confidence.

7.2.2. Analysis of marginal response invariance. Figure 7.2 shows the observed values and posterior distributions of Δ_{stop} (x -axes) and $\Delta_{fricative}$ (y -axes). As in onset position, for every subject, the observed values and posterior distributions are positive, in many cases with fairly

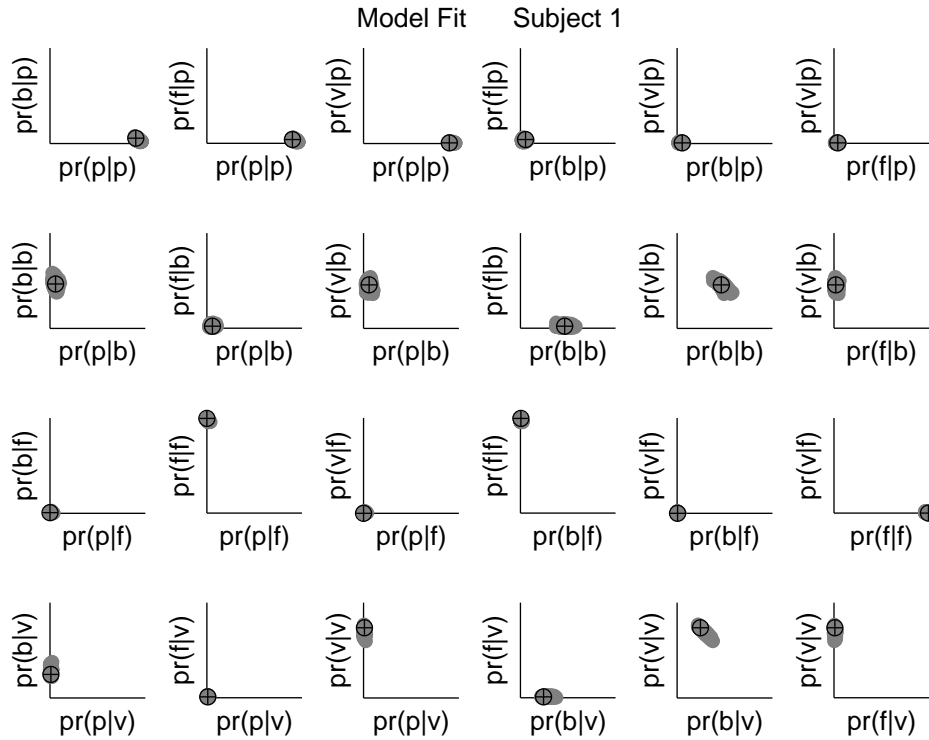


FIGURE 7.1. Scatter plots showing pairs of modeled identification-confusion probabilities (i.e., θ values; gray clouds) and observed identification-confusion proportions (black crosses) for subject one.

large values. Again, this indicates that subjects were more likely to respond ‘stop’ (‘p’ or ‘b’) when presented with [p] than when presented with [b], and that they were more likely to respond ‘fricative’ (‘f’ or ‘v’) when presented with [f] than when presented with [v]. The positive shift in Δ values for both levels of the manner dimension cannot be accounted for by a failure of decisional separability. Rather, this pattern of results is consistent with the presence of decisional separability and the failure of perceptual separability such that the manner distinction is more salient for voiceless [p] and [f] than it is for voiced [b] and [v]. The group level analysis reflects the individual level faithfully.

All eight subjects exhibit similar, and relatively small, levels of variability in their posterior Δ distributions. However, there are relatively large differences in the relative positive shift of Δ_{stop} and $\Delta_{fricative}$ between subjects. Three subjects – S1, S2, and S3 – exhibit similar degrees of positive

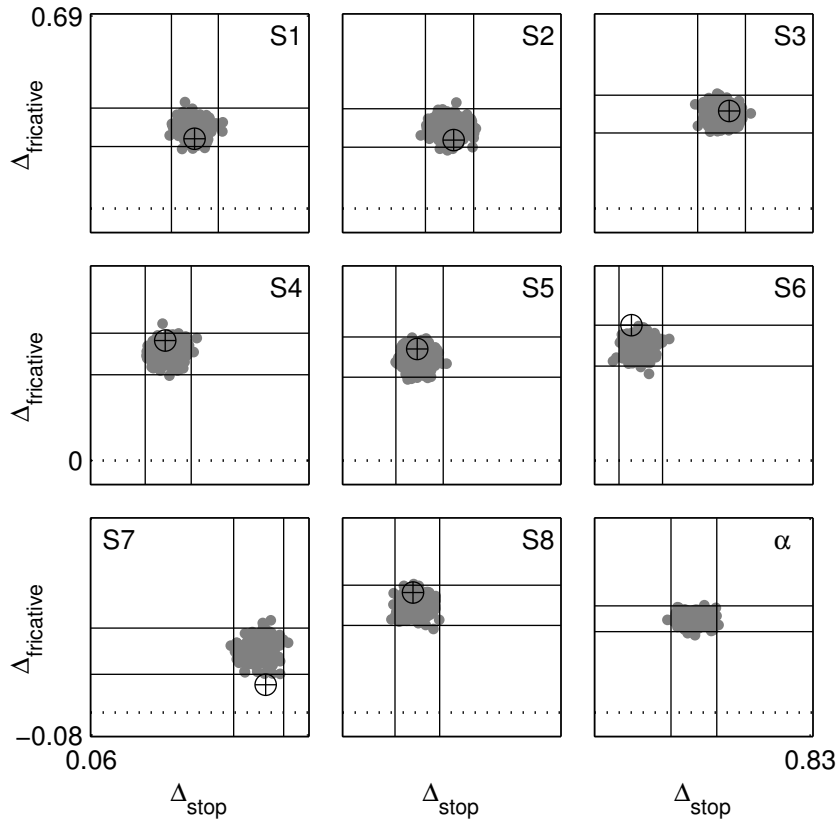


FIGURE 7.2. Scatter plots showing Δ_{stop} and $\Delta_{fricative}$ (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of sampled posterior Δ values; observed statistics are indicated by black circled crosses. The x -axes indicate values of Δ_{stop} ; the y -axes indicate values of $\Delta_{fricative}$. The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

shift for both, four subjects – S2, S4, S5, and S8 – exhibit a larger shift for $\Delta_{fricative}$ than for Δ_{stop} , while the remaining subject – S7 – exhibits the opposite pattern. The group level analysis shows fairly similar shifts for each level on the manner dimension.

S7's observed $\Delta_{fricative}$ value is outside the corresponding 99% highest probability density region. This is due to the hierarchical structure of the model. S7's data differ from the other seven subjects (again, by exhibiting a larger shift for Δ_{stop} than for $\Delta_{fricative}$), but the group level

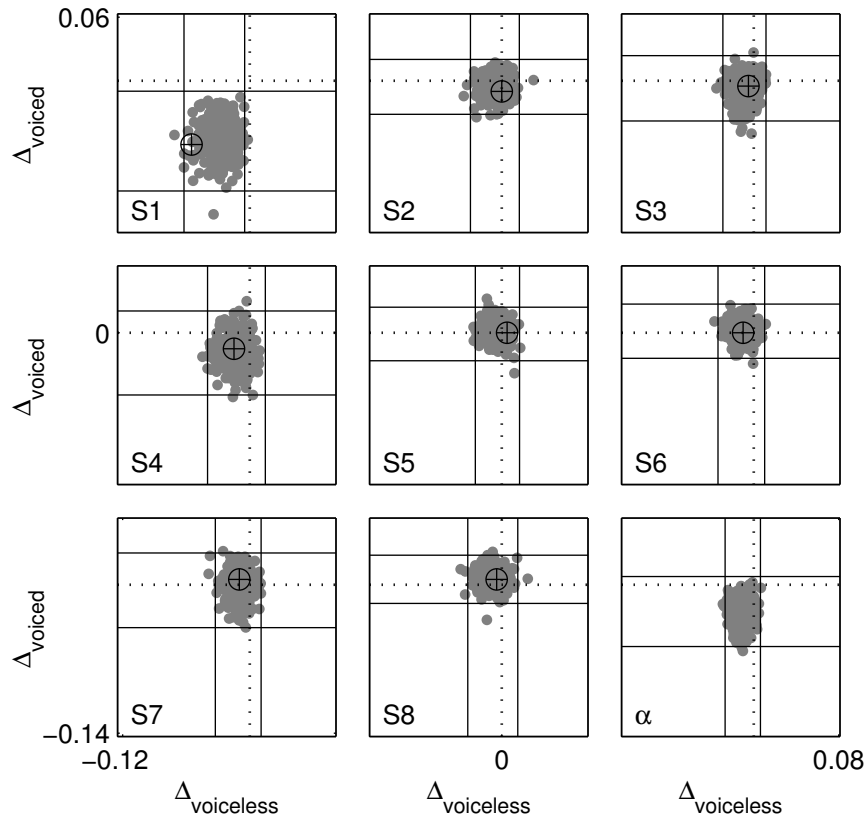


FIGURE 7.3. Scatter plots showing $\Delta_{\text{voiceless}}$ and Δ_{voiced} (equation 4) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of sampled posterior Δ values; observed statistics are indicated by black circled crosses. The x -axes indicate values of $\Delta_{\text{voiceless}}$; the y -axes indicate values of Δ_{voiced} . The bottom right panel (labeled α) shows these Δ statistics at the group level. Solid lines indicate limits on marginal 99% highest probability density regions; dotted lines indicate the location of zero.

structure in the model constrains the posterior distributions of Δ at the individual level, in this case by ‘pulling’ S7’s posterior $\Delta_{\text{fricative}}$ distribution toward more extreme positive values than it might otherwise consist of.

Figure 7.3 shows the observed values and posterior distributions of $\Delta_{\text{voiceless}}$ (x -axes) and Δ_{voiced} (y -axes). For all but S1, the highest probability density regions include zero for both statistics, indicating that marginal response invariance holds at both levels on the voicing dimension.

In most cases, the observed values for both statistics are very near zero, the distributions consist of roughly equal proportions of positive and negative values for Δ_{voiced} , and there is a consistent, though small, shift toward negative modeled $\Delta_{voiceless}$ values. For S1, both statistics (observed and modeled) are shifted substantially toward negative values. Thus, for all eight subjects, it is reasonable to posit that decisional separability holds, and for all but S1, that perceptual separability holds at both levels on the voicing dimension. S1's results indicate that the voicing distinction between [f] and [v] is more perceptually salient than that between [p] and [b].

The group level analysis indicates that marginal response invariance holds for both levels, though the posterior distributions for both statistics are shifted heavily toward negative values, perhaps in part because of S1's relatively extreme values.

As above, the variances in the posterior distributions of Δ_{voiced} and $\Delta_{voiceless}$ are much smaller than the variances in the posterior distributions of Δ_{stop} or $\Delta_{fricative}$.

7.2.3. Analysis of report independence. Figure 7.4 shows observed values and posterior distributions of π_p (x -axes) and π_b (y -axes). For all eight subjects, the 99% highest probability density regions for both statistics include zero. For π_p , though, the posterior distributions all consist almost entirely of positive values. On the other hand, the observed and modeled values of π_b are a bit more balanced on each side of zero, though there is some variability between subjects with regard to the occasional imbalance (e.g., S1's shift toward negative values as compared to the positive shift of S5 and S6).

The group level analysis also suggests that report independence holds for [b], though it diverges from the individual level analyses by indicating that report independence fails for [p]. As in the individual level analyses, the vast majority of group level π_p values are positive.

Figure 7.5 shows the observed values and posterior distributions of π_f (x -axes) and π_v (y -axes). The pattern here resembles the pattern observed with [p] and [b] in that, for all subjects, the posterior distributions of π_f (the voiceless member of this pair of consonants) consist almost entirely of positive values, whereas the posterior distributions of π_v (the voiced member of the pair) are more balanced on either side of zero. There appears to be a more consistent, though still small, shift toward positive values for π_v (compared to the distributions of π_b), and no subject has a posterior distribution shifted substantially toward negative values. For all but two subjects – S1 and S2 – the

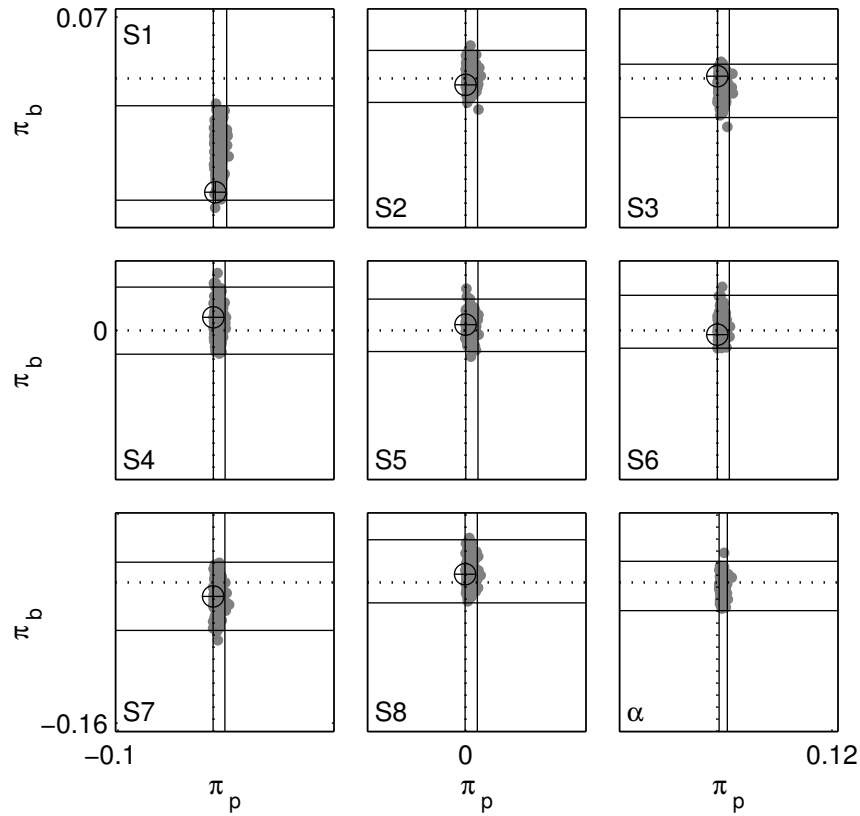


FIGURE 7.4. Scatter plots showing π_p and π_b (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for stop stimuli [p] and [b]. The x -axes indicate the value of π_p ; the y -axes indicate the value of π_b . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

highest probability density regions for π_f include zero, suggesting that report independence tends to hold for [f]. For all subjects, report independence clearly holds for [v].

The group level analysis indicates that report independence holds for [v] and fails for [f]. As was the case with [p] and [b] above, although the discrete inferences drawn from highest probability density regions differ in the individual and group level analyses, the posterior distributions of π_f are quite similar at each level of analysis.

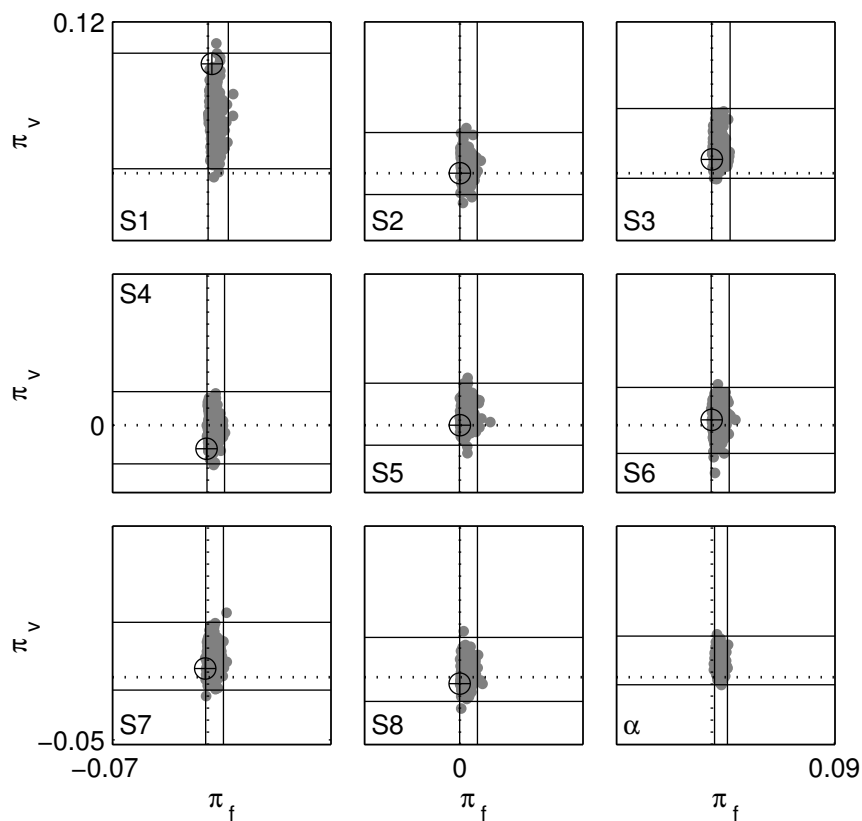


FIGURE 7.5. Scatter plots showing π_f and π_v (equation 5) for each subject and for the group as a whole. The gray clouds show a subset (1/10) of the posterior sampled values of π for fricative stimuli [f] and [v]. The x -axes indicate the value of π_f ; the y -axes indicate the value of π_v . Black circled crosses indicate observed values of π . The bottom right panel (labeled α) shows π for the group level. Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

7.2.4. Prediction of future data. Figure 7.6 shows predicted Δ and π values. Given the data observed, the model predicts that Δ_{stop} and $\Delta_{fricative}$ will be positive and of relatively large magnitude, indicating that marginal response invariance will fail. On the other hand, $\Delta_{voiceless}$ and Δ_{voiced} are predicted to be near zero, the latter is predicted to skew mildly towards negative values. The range of predicted values is much smaller for $\Delta_{voiceless}$ than for Δ_{voiced} , and smaller in general for the voicing dimension than for the manner dimension.

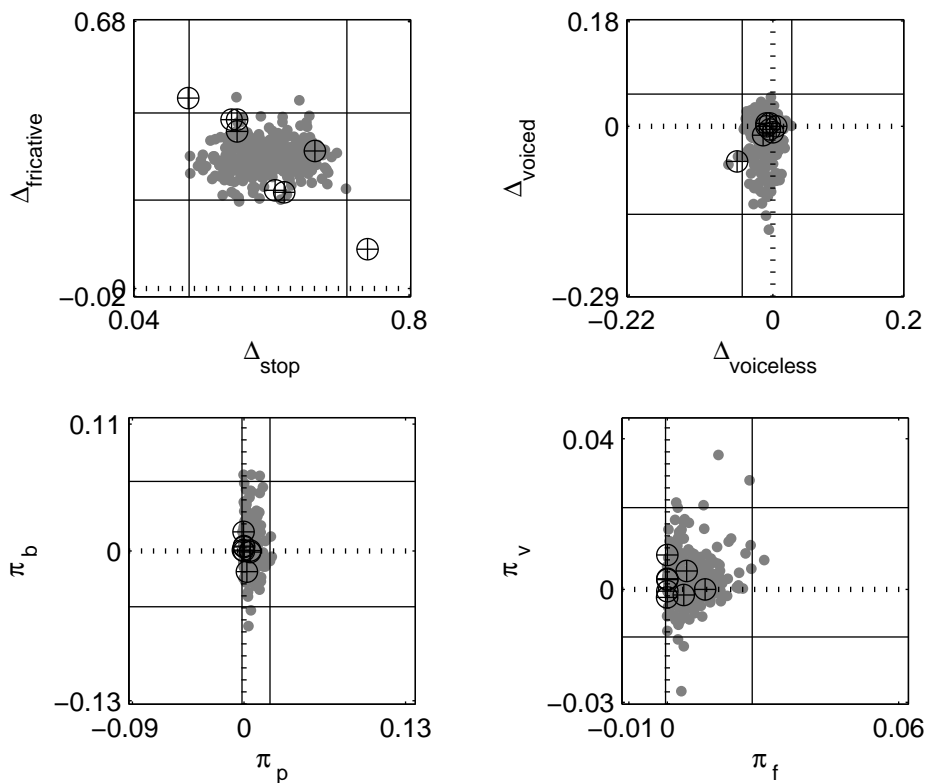


FIGURE 7.6. Scatter plots showing predicted and observed Δ and π statistics. The gray clouds show a subset (1/10) of the predicted values of Δ and π generated from the posterior distribution of the group level α ‘counts’. Black circled crosses indicate observed values of Δ and π . Solid lines indicate 99% marginal highest probability density regions; dotted lines indicate the location of zero.

Report independence is predicted to hold for all four consonants, with primarily positive values and small ranges of values for π_f and π_p and roughly equal proportions of positive and negative values and larger ranges of values for π_v and π_b .

7.2.5. Summary of results. Table 7.1 provides a summary of the analyses of marginal response invariance and report independence for manner of articulation and voicing in coda position. Overall accuracy was very high in this experiment, ranging from 70% to 88% correct responses.

All eight subjects exhibit failure of marginal response invariance for both ‘stop’ and ‘fricative’ responses. In each case, the failure is due to increased accuracy of manner identification for voiceless

[p] and [f] relative to voiced [b] and [v]. This strongly suggests an underlying failure of perceptual separability due to changes in the perceptual salience of manner across levels of voicing. Failure of decisional separability cannot account for the observed patterns.

On the other hand, marginal response invariance holds for all but one subject for ‘voiceless’ and ‘voiced’ responses, suggesting that voicing is both perceptually and decisionally separable from manner. For the one subject for whom marginal response invariance fails on the voicing dimension (S1), the pattern of results is consistent with the presence of decisional separability and failure of perceptual separability due to a change in salience of the voicing distinction across levels of manner. The group level analyses exhibit the same pattern of results as the individual level analyses.

Report independence holds in all but a small number of cases, consistent with both decisional separability and perceptual independence holding for manner and voicing in coda position. Because the analyses of marginal response invariance suggest that decisional separability holds for both dimensions, the small number of failures of report independence suggest occasional failures of perceptual independence.

7.3. Discussion

The patterns of interaction between manner of articulation and voicing vary consistently across onset and coda position. In both prosodic contexts, marginal response invariance fails for both ‘stop’ and ‘fricative’ responses. These failures always take the same form, namely large positive observed and modeled values of Δ_{stop} and $\Delta_{fricative}$. This pattern is consistent with an underlying model in which decisional separability holds and perceptual separability fails due to a change in salience of manner across levels of voicing.

Similarly, marginal response invariance holds for ‘voiceless’ responses in both onset and coda position, suggesting that voicing is decisionally separable from manner. On the other hand, whereas marginal response invariance tends to fail for ‘voiced’ responses in onset position, it tends to hold in coda position.

Report independence exhibits a similar pattern of differences across prosodic positions. For [p], [b], and [v], report independence holds for most (or all) subjects in both onset and coda position, and the posterior distributions of π_p are consistently shifted toward positive values for all subjects

TABLE 7.1. Summary of the results of analyses of marginal response invariance (Δ) and report independence (π). $p(C)$ = proportion correct responses. *stop* = [p] or [b]; *fric.* = fricative ([f] or [v]); *-voi* = voiceless ([p] or [f]); *+voi* = voiced ([b] or [v]). \checkmark indicates that marginal response invariance or report independence holds (i.e., that zero is within the 99% highest probability density region) for a given subject for that level and dimension (Δ) or stimulus (π).

Sub.	$p(C)$	Δ_{stop}	$\Delta_{fric.}$	Δ_{-voi}	Δ_{+voi}	π_p	π_b	π_f	π_v
1	0.87					\checkmark	\checkmark		
2	0.88			\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
3	0.78			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
4	0.85			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
5	0.80			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
6	0.70			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
7	0.80			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
8	0.81			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
α				\checkmark	\checkmark		\checkmark		\checkmark

in both syllable positions. On the other hand, report independence tends to fail for [f] in onset position and tends to hold for [f] in coda position.

As in the two experiments addressing place and voicing reported above, multiple subjects participated in both the onset and coda experiments probing manner and voicing. Table 7.2 shows a comparison of these subjects' results.

With regard to decision making, voicing and manner seem to generalize across one another, whereas with regard to perception, voicing seems to generalize (at least partially) across manner, but manner does not generalize across voicing. The degree of this generalization depends on syllable position such that it is only partial in onset position (i.e., at the voiceless level) while it occurs for both levels of voicing in coda position. By way of contrast, manner fails to generalize across voicing in either onset or coda position. In both syllable positions, manner is more perceptually salient for the voiceless pair [p] and [f] than it is for the voiced pair [b] and [v].

TABLE 7.2. Summary of the differences between onset and coda stop interactions.
 +/– indicate the same or different results of tests of marginal response invariance
 (Δ) and report independence (π) across experiments, respectively.

Sub.	Δ_{stop}	$\Delta_{fric.}$	Δ_{-voi}	Δ_{+voi}	π_p	π_b	π_f	π_v
1	+	+	–	+	+	+	+	+
2	+	+	+	–	+	+	+	+
3	+	+	+	–	+	+	–	+
4	+	+	+	–	+	+	–	+
5	+	+	+	–	+	+	–	+
8	+	+	+	–	+	–	–	+
α	+	+	+	–	+	+	–	+

Patterns of perceptual dependence between processing channels may vary across syllable position as well, at least in the voiceless fricative [f]. Report independence consistently fails for [f] in onset position, suggesting failure of perceptual independence, whereas it tends to hold for [f] in coda position. For the other three consonants – [p], [b], and [v] – report independence holds in both syllable positions.

The differences in patterns across prosodic positions lend weak support to the hypothesis that coda position induces, or allows for, greater perceptual feature generalization than does onset position. However, to the extent that this effect is present, it seems to be fairly weak. While the differences in marginal response invariance for place and voicing involved a level on each dimension (‘alveolar’ and ‘voiced’), the failures of marginal response invariance on the manner dimension were consistent across syllable positions; the increase in presence of marginal response invariance with respect to manner and voicing is restricted to the voiced level only.

Conclusion

Speech has multidimensional structure on at least two linguistically relevant levels – the acoustic-phonetic, and the phonological. Although each has been the focus of a number of studies of speech perception, a variety of theoretical and methodological limitations have left important questions unasked or unanswered. Among these limitations are conflation of logically distinct levels of processing and strong *a priori* assumptions about the relevant acoustic-phonetic cues to various phonological contrasts. The present work has sought to address these two limitations in an investigation of the integration of phonological information in speech perception.

Consider, again, the perception of an utterance containing the voiceless, bilabial stop [p]. As implied by this description, information from (at least) three linguistically relevant dimensions must be integrated – voicing, place of articulation, and manner of articulation. Depending on a large number of external and internal factors (e.g., crowd noise and speaking rate, respectively), a given token of [p] will be more or less ambiguous. It will be perceived to a greater or lesser degree as [b]-like, [t]-like, and [f]-like, which is to say that it will be perceived to a greater or lesser degree as voiced instead of voiceless, alveolar instead of labial, and fricative-like instead of stop-like.

The present work employed General Recognition Theory and a Bayesian non-parametric statistical model in order to characterize and investigate the integration of information from, on the one hand, voicing and place of articulation, and on the other, voicing and manner of articulation. Each of these phonological subspaces was analyzed in both syllable onset and syllable coda positions.

8.1. General Recognition Theory and a multinomial-Dirichlet statistical model

8.1.1. General Recognition Theory. GRT characterizes information integration in terms of independence (or lack thereof) between processing dimensions and provides definitions for three logically distinct forms of independence. A multidimensional generalization of Signal Detection Theory, GRT relies on two assumptions: randomly distributed perceptual effects and deterministic decisions concerning these perceptual effects. The simplest experimental protocol that will allow

for all three forms of independence to be investigated simultaneously consists of identification of four stimuli defined as the factorial combination of two levels on each of two dimensions (e.g., labial and alveolar levels on the place of articulation dimension and voiced and voiceless on the voicing dimension, as in experiments one [Chapter 4] and two [Chapter 5] above).

Perceptual independence is defined at the level of the individual perceptual distribution. If stochastic independence holds within a given perceptual distribution, perceptual independence holds for the corresponding stimulus. Perceptual separability is defined across perceptual distributions. If the perceptual effects on one dimension (e.g., place of articulation) are identical across the other dimension (e.g., voicing), the place of articulation is perceptually separable from voicing. Finally, decisional separability holds if decisions about one dimension do not depend on decisions or perception on the other dimension.

Of course, the perceptual distributions and decision bounds of GRT are unobservable. Ashby and Townsend (1986) showed that if certain combinations of forms of independence hold, then certain empirical measures of interaction will hold. On the one hand, if perceptual separability and decisional separability hold, this implies that marginal response invariance will hold (see equations 1 and 2). On the other hand, if perceptual independence and decisional separability hold, this implies that report independence will hold (see equation 3).

8.1.2. A non-parametric statistical model. In order to quantify the uncertainty of our observations, a hierarchical Bayesian non-parametric model was developed. The model relies on two very general assumptions, namely that, for each stimulus, identification-confusion frequencies are distributed as multinomial random variables at the individual subject level, and that, at the group level, the probabilities underlying the multinomial random variables are distributed as Dirichlet random variables.

The BRugs software package was used to sample from the joint posterior distribution of the modeled (individual level) identification-confusion probabilities and (group level) identification-confusion counts. The resulting posterior distributions were then transformed to obtain distributions of statistics measuring marginal response invariance (Δ , defined in equation 4, section 3.1, and repeated below) and report independence (π , defined in equation 5, section 3.1, and repeated below). The observed values and modeled distributions of Δ and π enable inferences regarding both whether marginal response invariance and report independence hold and, if they do not hold, how and to

what degree they fail. The posterior distributions from the Bayesian model also enable quantitative predictions regarding future data.

$$\begin{aligned}\Delta &= p(a_i|A_i, B_1) - p(a_i|A_i, B_2) \\ e.g., &= p(\text{'labial'}|[p]) - p(\text{'labial'}|[b]) \\ \pi &= p(a_i, b_j|A_i, B_j) - \left[p(a_i|A_i, B_j) \times p(b_j|A_i, B_j) \right] \\ e.g., &= p(\text{'labial, voiceless'}|[p]) - \left[p(\text{'labial'}|[p]) \times p(\text{'voiceless'}|[p]) \right]\end{aligned}$$

Thus, the non-parametric Bayesian multinomial-Dirichlet model relies on very general assumptions about the nature of the data, explicitly models group-level patterns, and provides valuable gradient information about the measures of interaction Δ and π . By way of contrast, frequentist methods for testing marginal response invariance (e.g., z -tests) and report independence (e.g., χ^2 tests of independence) rely on strong assumptions about the statistics of interest (rather than the raw data), do not (currently) model group level structure, and, at least with regard to χ^2 tests of report independence, provide substantially less information about how and to what degree report independence fails.

8.2. Recapitulation of experimental results

8.2.1. Experiment One. These methods were first applied to an investigation of the relationships between place of articulation and voicing in syllable onset stop consonants. The stimuli consisted of four tokens of each of the nonsense syllables [pa], [ba], [ta], and [da], which represent a factorial combination of the labial ([p] and [b]) and alveolar ([t] and [d]) levels on the place of articulation dimension and the voiceless ([p] and [t]) and voiced ([b] and [d]) levels on the voicing dimension. This set of consonants has been probed before (e.g., Oden & Massaro, 1978; Eimas et al., 1978), though these earlier studies have serious limitations. The methods employed here were developed, in part, to address these limitations.

The results indicate that, both at individual subject and group levels, marginal response invariance tends to hold for ‘labial’ and ‘voiceless’ responses, tends to fail for ‘voiced’ responses, and fails across the board for ‘alveolar’ responses. Taken together, the patterns observed are consistent with an underlying model in which decisional separability holds for both dimensions, and in which perceptual separability fails due to a greater perceptual salience of [d] on each dimension (i.e., [d] is

more salient from [b], relative to the salience of [t] from [p], and is more salient from [t], relative to the salience of [b] from [p]). The results also indicate that report independence tends to hold for all four consonants.

The failures of perceptual separability indicate that the features place of articulation and voicing do not generalize across one another perceptually. On the other hand, the apparent presence of decisional separability suggests that at the post-perceptual, decisional level, these features do generalize across one another. Given that the analyses of marginal response invariance suggest that decisional separability holds, the fact that report independence tends to hold suggests that perceptual independence tends to hold, and that failures of perceptual independence have caused the small number of failures of report independence. Thus, at the level of the perceptual distribution, place and voicing seem to be independent.

These results might help explain apparently contradictory results reported in earlier work. Oden and Massaro (1978) argue that their results support independence between place and voicing, whereas Eimas et al. (1978) argue that these dimensions interact. To the extent that the independence claimed by Oden and Massaro (1978) corresponds to the GRT notion of perceptual independence, and to the extent that the interactions claimed by Eimas et al. (1978) are due to failures of perceptual separability, the present work provides a reasonable explanation for both.

8.2.2. Experiment Two. The same methods were next applied to an investigation of the relationships between place and voicing in syllable coda position. In a simple, if naive, phonological framework we might hypothesize that the relationships between two (or more) features are the same regardless of where they are implemented within a larger structure, such as a syllable (or word, phrase, sentence, etc...). Basic phonetic knowledge tells us that, in speech production, the acoustic-phonetic cues to place and voicing are very different in syllable onset and syllable coda positions. However, the mere fact that acoustic-phonetic cues differ across syllable position does not guarantee that patterns of separability and independence will, too.

The results indicate that marginal response invariance may hold or fail for ‘labial’ responses, and tends to hold for ‘alveolar’, ‘voiceless’, and ‘voiced’ responses; at the group level, marginal response invariance holds for each level on each dimension. Again, these results are consistent with the presence of decisional separability on both dimensions. However, unlike the analyses of these features in onset position, the coda position results are also consistent with the presence of perceptual

separability in both dimension (for most subjects). The report independence results suggest that perceptual independence also tends to hold for place and voicing in coda position.

With respect to place and voicing, the main difference between onset and coda position seems to be a difference in the degree to which features generalize across one another perceptually. More specifically, it seems to be the case that place and voicing are more likely to generalize across one another in coda position than in onset position. Though it is beyond the scope of this work to draw firm conclusions regarding the cause of the difference between onset and coda position, the substantial differences in the acoustic-phonetic cues to voicing and place (e.g., stop release burst and VOT in onset position, lack of same in coda position) and the differences in how they interact (e.g., destruction of formants and formant transitions due to ‘cut-back’ in voiceless stops in onset position, lack of same in coda position) suggest themselves as plausible candidates.

8.2.3. Experiment Three. The phonological subspace probed in experiments one and two was chosen in part because earlier work had probed the same dimensions using the same consonants and in part because it is a convenient subspace of features that fits well in the standard two-by-two GRT structure. This place by voicing space also happens to overlap with another convenient phonological subspace in English, namely the voicing by manner of articulation space consisting of [p], [b], [f], and [v] (see Figure 6.1). These consonants consist of the factorial combination of the stop ([p] and [b]) and fricative [f] and [v] levels on the manner dimensions and the voiceless ([p] and [f]) and voiced ([b] and [v]) levels on the voicing dimension.

The results indicate that, both at individual subject and group levels, marginal response invariance fails for ‘fricative’, ‘stop’ and ‘voiced’ responses and holds for ‘voiceless’ responses. The pattern of failures of marginal response invariance on the manner dimension are consistent with an underlying model in which decisional separability holds and perceptual separability fails due to a roughly symmetrically greater salience between [p] and [f] than that between [b] and [v]. The pattern of results on the voicing dimension are also consistent with the presence of decisional separability and the failure of perceptual separability, though the failure of perceptual separability seems to be restricted to the voiced level such that [v] is more distinct from [f] than [b] is from [p].

At the individual subject level, report independence tends to hold for [b], [v], and [p], though for every subject it nearly fails for [p]; report independence tends to fail for [f]. At the group level, report independence holds for [b] and [v] and fails for [p] and [f]. Given that the analyses of marginal

response invariance suggest that decisional separability holds, the failures of report independence with respect to [f] (and the near-failures with respect to [p]) seem to be due to failures of perceptual independence such that there is a larger proportion of the perceptual distributions falling in the correct response regions than in the adjacent partially correct regions (relative to a case in which perceptual independence holds).

These results indicate that, as with place and voicing, manner and voicing seem to generalize across one another in decision making but do not generalize across one another perceptually. With respect to place and voicing, the labial stops [p] and [b] appeared to be perceptually separable (i.e., labial perceptual effects generalize across levels of voicing). With respect to manner and voicing, the same labial stops are not perceptually separable (i.e., stop manner does not generalize across voicing levels). The labial fricatives [f] and [v] exhibit a similar pattern (i.e., fricative manner does not generalize across voicing either). The apparent failures of perceptual independence suggest that the channels processing manner interact for the voiceless [f] and, possibly, [p], but not for voiced [b] and [v].

8.2.4. Experiment Four. The same methods were applied a final time to the relationships between manner and voicing in syllable coda position (i.e., [p], [b], [f] and [v] at the ends of nonsense syllables). Although there does not appear to be any previous work addressing these issues with respect to these consonants, experiments one and two suggest a hypothesis concerning the modulation of feature interactions by syllable structure. Specifically, if the pattern observed across experiments one and two is an instance of a more general phenomenon, we should expect a similar increase in perceptual generalization of manner and voicing in coda relative to onset position.

The results indicate that marginal response invariance fails for ‘stop’ and ‘fricative’ responses, again in such a way that is consistent with decisional separability and a failure of perceptual separability due to greater salience between [p] and [f] than between [b] and [v]. Marginal response invariance tends to hold for both ‘voiceless’ and ‘voiced’ responses in coda position, whereas it holds only for ‘voiceless’ responses in onset position. Report independence tends to hold for all four consonants in coda position, suggesting that perceptual independence tends to hold. Again, though, there are numerous instances of near-failures.

As with onset position, manner and voicing seem to generalize across one another in decision making and manner does not generalize perceptually across voicing in coda position. On the other

hand, voicing does seem to generalize perceptually across manner in coda position, whereas it does not in onset position. This pattern of results is consistent with the hypothesis that features tend to generalize across one another more in coda position than in onset position. Of course, consistency is a weak relationship, and the present work was not carried out with evaluation of this specific hypothesis in mind.

8.3. Possible Causes and Implications

Determining the causes and implications of the phenomena reported here is beyond the present work. Nonetheless, a brief discussion of some possible causes and implications of this work will situate it in the larger body of work on speech production and perception.

8.3.1. Possible causes. Generally speaking, there are two possible causes for the patterns of interaction and independence reported here. On the one hand, it may be that people’s (linguistic) perceptual systems accurately track fine-grained details in stimuli. So-called ‘exemplar’ models of perception making this assumption have been shown to account well for interesting perceptual data (e.g., Nosofsky & Zaki, 2002; Nosofsky & Stanton, 2005). To the extent that speech perception is driven by exemplar-based mechanisms, we might expect the patterns reported here to reflect the same patterns in acoustic-phonetic space. For example, this approach would predict an irregular place by voicing space for onset stops wherein [da] syllables reside relatively far away from [ta] and [ba] syllables, whereas place by voicing space should form a more regular shape for coda stops, with [ad] syllables forming one corner of a rectangular space. Similarly, we should expect distributions of both onset and coda [p] and [f] to exhibit less overlap than do analogous distributions of [b] and [v], and we should expect voicing to generalize completely across manner in coda labials but only incompletely in onset labials.

On the other hand, it is at least *a priori* plausible to suggest that people’s perceptual systems impose endogenous structure on stimuli. There is, in fact, some experimental work suggesting that this is the case, at least with respect to the integration of various cues to voicing in intervocalic stops (Kingston et al., 2008). To the extent that perceptual independence or interactions are due to structure imposed by the perceptual system, we should expect divergence between the patterns of relationships between dimensions in production and perception.

These possibilities point the way toward a program of research designed to tease them apart. Such a program is outlined briefly in the ‘future work’ section below.

8.3.2. Possible implications. The apparent interactions between place and voicing and between manner and voicing reported here could have a number of affects on ‘downstream’ (i.e., post-phonological) speech processing. First, the structure of phonological neighborhoods may reflect the differences in salience observed here. For example, lexical neighbors differing only in the voicing of onset *alveolar* stops (e.g., ‘ton’ and ‘done’) should be more distinct than analogous neighbors that differ only in the voicing of onset *labial* stops (e.g., ‘pun’ and ‘bun’); similar predictions can be made based on the voicing and manner interactions reported here. Sufficiently large differences in salience across contrasts may well modulate lexical neighborhood effects. If the bulk of a word’s (high-frequency) neighbors differ from that word only by more salient distinctions, we might reasonably expect neighborhood effects on processing that word to be weaker than similar effects on processing a word with perceptually less distinct neighbors.

Second, the interactions reported here may have potentially important implications for perceptual learning in various other listener groups (e.g., non-native listeners, people with age-related or noise-induced hearing loss, cochlear implant users). Depending on the cause(s) of perceptual feature generalization, we might expect perceptual learning to reflect the extent to which a given feature generalized perceptually across another feature. Suppose, for example, that perceptual interactions reflect interactions in production, as discussed immediately above. To the extent that acoustic-phonetic cues in two different languages map similarly onto phonological dimensions, we should expect non-native listeners of one to perceive the other in much the same way (i.e., we should expect features to interact, or not, to roughly the same extent and in the same way). If acoustic-phonetic cues in two languages map onto phonological dimensions differently, though, we should expect systematic differences in patterns of feature interactions when native listeners of one language perceive speech sounds in the other (e.g., Japanese native listeners perceiving English labial stops, as suggested by the differential use of acoustic-phonetic cues by non-native listeners in Oglesbee, 2008). If, on the other hand, we suppose that general characteristics of human auditory systems govern the relationships between phonological features in speech perception, we might reasonably expect similar patterns of feature interactions regardless of a listener’s native language.

Similar implications may be hypothesized for people with age-related or noise-induced hearing loss and cochlear implant users. If perceptual interactions reflect distributions of acoustic-phonetics cues, then to the extent that the utility of cues is reduced by hearing loss or the degraded signal provided by a cochlear implant, we should expect differences in the patterns of perceptual interactions observed. For example, if the acoustic-phonetic cues that made [d] so distinctive in experiment one are in a region of reduced audibility, or if these cues are not transduced effectively by a cochlear implant, we might expect listeners with hearing loss and cochlear implant users to exhibit a higher degree of feature generalization of voicing and place in onset position than did the native listeners described above. On the other hand, if it is reasonable to hypothesize that normally-functioning auditory systems might impose structure on speech input, then it seems just as reasonable to hypothesize that a system with a cochlear implant front end or one changed by hearing loss might do so, as well, distinct though the specific form of such structure is likely to be.

Training regimens may be able to exploit the degree of generalization exhibited by a group of listeners by reducing the amount of exposure required to learn a distinction. If voicing generalizes across place (as it seems to do in coda position), for example, non-native listeners or listeners with hearing loss may be able to learn to perceive a general voicing contrast without reference to specific places of articulation. Conversely, we might expect differences in learning and effective training regimes in cases in which generalization is not complete. If voicing does not generalize across place (as seems to be the case in onset position), listeners may need to learn to perceive specific combinations of voicing and place.

8.4. Limitations

There are, of course, numerous limitations to any experimental science. To state perhaps the most obvious example, the results obtained here only generalize to ‘natural’ speech perception to the extent that the experimental stimuli and task approximate ‘natural’ speech perception. The present work is no exception, and two limitations to this are worth discussing in some detail. These are the influence of the type of masking noise used and the stimulus properties.

8.4.1. Noise type. As described in Chapter 4, a ‘speech shaped’ noise was used to mask the speech stimuli. Because errors are crucial to the application of a model like GRT, it was necessary to use some kind of masking noise in order to ensure that experimental subjects responded erroneously

some portion of the time.¹ Speech shaped noise was chosen here for two reasons. First, it enables fairly strict control over the spectral and temporal characteristics of the masker, and second, it masks all three of the features under investigation to a similar degree. White noise would have provided much more control over the acoustic characteristics of the masker, but, based on pilot work and other research (G. A. Miller & Nicely, 1955), would have masked the acoustic-phonetic cues for place and manner far more effectively than the cues for voicing. Multi-talker babble, on the other hand, has complementary benefits and problems; it provides very little fine grained control over the spectral or temporal properties of masker but masks place, manner, and voicing roughly equally well.

It is unknown to what extent the ‘speech-shaped’ noise employed here is responsible for the observed patterns of (failure of) marginal response invariance and report independence. There are theoretical and empirical aspects to this issue. With regard to the theoretical aspects, there is no obvious reason that decision making should vary as a function of the noise type employed, whereas it is possible that different noise types could differentially induce different patterns of perceptual interaction or independence. If, say, multi-talker babble masks manner cues better for voiced than for voiceless segments while white noise masks manner cues equally well regardless of voicing, we would expect differences in observed (failures of) perceptual separability and, thereby, marginal response invariance. Similarly, if we interpret failure of perceptual independence as correlated noise across processing channels, it is plausible that different noise types could produce different levels or types of interaction between channels, and thereby different patterns of (failure of) report independence. Ultimately, of course, it is an empirical issue whether or not different noise types induce different types of perceptual interactions.

8.4.2. Stimulus properties. As described above, four naturally produced stimuli per category were employed. This had two motivations. First, it enabled us to avoid making strong *a priori* assumptions about the acoustic-phonetic cues relevant to place, manner, or voicing. Second, it allowed us to obtain a reasonable amount of data per stimulus while providing enough within category variability to prevent subjects from identifying stimuli by attending to acoustic-phonetic that are exploitable in a psychophysical experiment but linguistically irrelevant.

¹ Errors may, of course, be induced without masking noise. For example, the stimuli could have been presented at near-threshold levels, or the task could have included extreme time pressure.

Avoiding strong assumptions in this way, however, severely limits the strength of any inferences drawn regarding the causes of the observed perceptual interactions. At best, correlations between particular acoustic-phonetic properties of the stimuli and the observed patterns of identification and confusion might be investigated with the present data and analyses. While this might provide interesting information, it would be of more value, in the long run, to design studies with the intention of understanding how particular combinations of acoustic-phonetic cues lead directly to perceptual independence or interactions.

Although the four stimuli per category do seem to have provided the desired within-category variability, they do not provide much general information about the nature of distributions of acoustic-phonetic cues on phonological dimensions in speech production. This information would be crucial in any investigation of the causes of the observed patterns of interaction. It may be that the observed patterns indicate that a native speaker's perceptual system closely tracks natural variation in production; the locations and shapes of perceptual distributions may accurately reflect production distributions.

On the other hand, it may be that native speakers perceptual systems transforms speech input. If so, we should expect systematic differences between production distributions and perceptual distributions. This is an empirical matter. A larger number of stimuli could provide a richer sample of production distributions, enabling such tests to be carried out. A full characterization of production distributions could also inform active manipulations of acoustic-phonetic cues in the search for the causes of perceptual interactions.

8.5. Future work

There are numerous ways in which the present work can be extended. First, the data analyzed and discussed here can be subjected to a variety of other analyses. Second, as noted, the data addressed here are a subset of the data originally collected; the rest of the data can be analyzed in the same way the presented data can. Third there are numerous separate projects related to various aspects of the present work. These three types of future work will be discussed in turn.

8.5.1. Other analyses. Although emphasis was placed on non-parametric methods here, there are two different ways in which making stronger assumptions and employing parametric techniques can produce additional useful and interesting results in the GRT framework. The first is the use

of Signal Detection Theory to probe marginal and response conditioned d' and β parameters. The second is fully parameterized model fitting.

The use of d' and β to investigate the multidimensional structure of identification data predates the development of GRT proper (e.g., Townsend et al., 1981, 1984), though the methods were later more fully developed with respect to the GRT notions of independence and interaction (Kadlec & Townsend, 1992a, 1992b). To mention just one of the important results relating SDT to GRT, Kadlec and Townsend (1992a, 1992b) prove that perceptual separability implies equality of d' on one dimension across levels of the other. For example, if voicing were separable from place (as it seems to be in coda consonants), we would expect d'_{pb} to be equal to d'_{td} . Methods for relating β to decisional separability have also been developed, as have methods for relating response-conditioned d' and β to perceptual independence. These methods could be readily applied both to observed identification-confusion proportions and to modeled identification-confusion probabilities to obtain posterior distributions of marginal and response-conditioned d' s and β s.

Fully parameterized GRT models can also be fit to identification-confusion data. For example, the mean vectors and covariance matrices of bivariate Gaussian distributions and the intercepts and slopes of linear decision bounds can be used to calculate predicted identification-confusion probabilities, which can then be used to evaluate the fit of a given set of parameters to a given data set. This kind of model fitting can be employed in maximum likelihood algorithms (i.e., algorithms that find the single best fitting set of parameters) or in a Bayesian framework in which prior probability distributions are placed on the parameters and the analysis produces posterior distributions on same. Perceptual and decisional interactions are then evaluated by investigating the (posterior distributions of) fitted parameters. The major drawback to the use of fully parameterized models is the substantial increase in computational power and time needed to carry out analyses. The non-parametric Bayesian analyses presented here required only a few minutes time to carry out, whereas maximum likelihood methods with Gaussian GRT models can take days or even weeks to complete the required computations.

8.5.2. Other data. Only equal base rate identification-confusion data were analyzed here. These data were collected along with data in four unequal base rate conditions (i.e., experimental blocks in which the relative presentation frequencies of the levels on one dimension were shifted from

1-to-1 to 4-to-1). Response times were also collected (subjects were instructed to be as accurate and as fast as they could be in responding on each trial).

The identification-confusion data from the shifted base rate conditions can, of course, be subjected to any analytic technique that the equal base rate identification-confusion data can be subjected to. In addition to providing additional tests of perceptual and decisional independence, this extra data could, in theory, enable tests of hypotheses regarding the effect of shifted base rates. For example, Balakrishnan (1999) argues that shifting the base rates of stimuli changes the variance of perceptual distributions while leaving decision criteria unchanged.

The response times collected provide a variety of interesting opportunities for investigating the integration of phonological information in speech perception. For example, the non-parametric measures of interaction used here – marginal response invariance and report independence – can be given temporal definitions and investigated with appropriate response time data.² Of course, parametric models may also be used to analyze response time data. Given the complexity of GRT data, the most promising avenue here seems to be the Linear Ballistic Accumulator model of choice response time (Brown & Heathcote, 2008), which simplifies the diffusion model that has been used extensively with two-choice data (e.g., Ratcliff, 2002). The major advantage that the Linear Ballistic Accumulator has over other related models is the fact that it generalizes to any number of responses quite easily, whereas the more complex diffusion models do not.

8.5.3. Related separate projects.

8.5.3.1. *Other parts of phonological space.* Perhaps the most obvious extension of the present work would be to other phonological subspaces. This could be done in a number of ways. The place by voicing experiments presented in Chapters 4 and 5 could be extended to include the velar place of articulation (i.e., [k] and [g]), defining three levels on the place dimension. Manner and place could be probed for either voiced or voiceless segments (or both). Of course, other features could also be included; for example, English nasal and non-nasal consonants [m], [n], [b], and [d] fit the standard GRT structure nicely. Vowel systems (or portions thereof) are also amenable to this line of research.

8.5.3.2. *Other languages.* It has long been known that languages differ in their instantiation of the same (or at least similar) phonological contrasts (e.g., Lisker & Abramson, 1964). To the extent

² This work is currently being carried out in the Townsend lab.

that particular combinations of acoustic-phonetic cues induce particular perceptual (or decisional) interactions (or independence), we should expect systematic variation across languages. Closely related is the nature of native vs. non-native speech perception. To the extent that listeners are attuned to the particular combinations of acoustic-phonetic cues and phonological contrasts of their native language, we should expect interesting differences to arise when identifying native vs. non-native sounds.

8.5.3.3. *Production distributions.* As noted in the ‘limitations’ section above, the four stimuli per category do not provide much information about the distribution of acoustic-phonetic cues in English in general. Larger numbers of tokens should provide representative samples of phonological categories, and the definitions of perceptual separability and independence are readily adapted to multivariate statistical modeling of speech production data. For example, multiple [p], [b], [t], and [d] tokens from multiple talkers could be measured on multiple acoustic-phonetic dimensions, which could then be projected onto bivariate Gaussian distributions in a two dimensional phonological space. Within and between category forms of independence could then be investigated in this production space.

While this would be interesting in its own right, it could also be directly, and interestingly, related to perceptual data gathered using some (or all) of the tokens from the production task. For example, the degree to which listeners track variation in production could be investigated in a reasonably straightforward manner. As discussed above, in the ‘possible causes’ section, if listeners perceive speech veridically, then modeled perceptual distributions should closely mirror production distributions. If, on the other hand, listeners’ perceptual systems transform the input, then perceptual distributions should differ systematically from production distributions.

Active manipulation of (distributions of) acoustic-phonetic cues could, in theory, aid in such investigations. Should it be observed, apparently veridical perception could be due either to the enormous amount of exposure adult listeners have with the natural distributions of speech sounds in their native language(s) or to particular combinations of acoustic-phonetic cues causing particular patterns of interaction and independence. The use of synthesized (or resynthesized) speech could enable disambiguation of these two possibilities. By creating artificial distributions of acoustic-phonetic

cues (i.e., distributions that differ systematically from the naturally observed distributions), the degree to which perceptual interactions are driven by particular combinations of acoustic-phonetic cues or the statistical properties of the distributions of such cues could be measured.

8.6. Closing remarks

Despite the fact that perceptual and decisional processes are not directly observable, inferences about perception and decision making can be drawn with the right tools. The present work employed one such tool – General Recognition Theory – in an effort to draw inferences about the relationships between phonological dimensions in speech perception. Although GRT is not a perfect tool (e.g., distinct underlying models predict the same patterns in identification data, as illustrated in Figures 3.3 and 3.4, section 3.3), it can provide us with valuable insights into the structure and processes that constitute perception. Of equal, if not greater, significance, mathematical models such as GRT enable us to ask precise, pointed questions. These questions would be difficult to conceive of, much less formulate, in the absence of the kind of theoretical framework provided by GRT and its ilk. Once formulated, these questions can serve as powerful guides in future work.

References

- Ashby, F. G. (1989). Stochastic general recognition theory. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models* (p. 435-457). Amsterdam: Elsevier.
- 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., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *91*(2), 154-179.
- Balakrishnan, J. D. (1999). Decision processes in discrimination: Fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1189-1206.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153-178.
- DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review*, *109*(4), 710-721.
- Eimas, P. D., Tartter, V. C., & Miller, J. L. (1981). Dependency relations during the processing of speech. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (p. 283-309). Lawrence Erlbaum Associates.
- Eimas, P. D., Tartter, V. C., Miller, J. L., & Keuthen, N. J. (1978). Asymmetric dependencies in processing phonetic features. *Perception & Psychophysics*, *23*(1), 12-20.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons, Inc.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of english fricatives. *Journal of the Acoustical Society of America*, *108*(3), 1252-1263.

- 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*, 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* (p. 181-227). New Jersey: Lawrence Erlbaum Associates.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419-454.
- Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, *36*, 28-54.
- Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *The Journal of the Acoustical Society of America*, *97*(2), 1261-1285.
- Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., & Bartels, C. (1997). Integrality in the perception of tongue root position and voice quality in vowels. *The Journal of the Acoustical Society of America*, *101*(3), 1696-1709.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.
- Macmillan, N. A., Kingston, J., Thorburn, R., Dickey, L. W., & Bartels, C. (1999). Integrality of nasalization and f1. ii. basic sensitivity and phonetic labeling measure distinct sensory and decision-related interactions. *The Journal of the Acoustical Society of America*, *106*(5), 2913-2932.
- Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (p. 147-180). New Jersey: Lawrence Erlbaum Associates.
- Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 288-301.
- Maddox, W. T., & Bohil, C. J. (1998a). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1459-1482.

- Maddox, W. T., & Bohil, C. J. (1998b). Overestimation of base-rate differences in complex perceptual categories. *Perception & Psychophysics*, *60*(4), 575-592.
- Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, *67*(3), 996-1013.
- Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *Journal of the Acoustical Society of America*, *27*(2), 338-352.
- Miller, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics*, *24*(2), 175-180.
- Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *Journal of the Acoustical Society of America*, *61*(3), 835-845.
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*(6), 505-512.
- Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, *18*(3), 347-373.
- Nearey, T. M. (1992). Context effects in a double-weak theory of speech perception. *Language and Speech*, *35*(1/2), 153-171.
- Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*(6), 3241-3254.
- Nearey, T. M. (2001). Phoneme-like units and speech perception. *Language and Cognitive Processes*, *16*(5/6), 673-681.
- Nearey, T. M. (2003). On the factorability of phonological units in speech perception. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology vi: Phonetic interpretation* (p. 197-221). Cambridge: Cambridge University Press.
- Nearey, T. M., & Hogan, J. T. (1986). Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual categorization curves. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (p. 141-161). FL Orlando: Academic Press.
- Nosofsky, R. M., & Stanton, R. D. (2005). Speeded classification in a probabilistic category structure: Contrasting exemplar-retrieval, decision-boundary, and prototype models. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 608-629.

- Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 924-940.
- Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*(3), 172-191.
- Oglesbee, E. N. (2008). *Multidimensional stop categorization in english, spanish, korean, japanese, and canadian french*. Unpublished doctoral dissertation, Indiana University.
- Olzak, L. A. (1986). Widely separated spatial frequencies: Mechanism interactions. *Vision Research*, *26*(7), 1143-1153.
- Olzak, L. A., & Wickens, T. D. (1997). Discrimination of complex patterns: Orientation information is integration across spatial scale; spatial-frequency and contrast information are not. *Perception*, *26*, 1101-1120.
- Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a brightness discrimination task. *Psychonomic Bulletin & Review*, *9*, 278-291.
- Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics*, *2*, 181-194.
- Silbert, N., & Jong, K. de. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America*, *123*(5), 2769-2779.
- Silbert, N., Townsend, J. T., & Lentz, J. J. (2007a). Independence in the perception of complex non-speech sounds. *Journal of the Acoustical Society of America*, *122*, 3062.
- Silbert, N., Townsend, J. T., & Lentz, J. J. (2007b). *Independence in the perception of complex non-speech sounds*. (Poster presented at the 40th Annual Meeting of the Society for Mathematical Psychology, University of California, Irvine.)
- Smits, R. (2001a). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1145-1162.
- Smits, R. (2001b). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics*, *63*(7), 1109-1139.
- Stevens, K. N. (1998). *Acoustic phonetics*. MA Cambridge: MIT Press.

- Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America*, *91*, 2979-3000.
- Team, R. D. C. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.
- Thomas, R. D. (2001a). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (p. 193-227). New Jersey: Lawrence Erlbaum Associates.
- Thomas, R. D. (2001b). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics*, *63*(4), 625-650.
- Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research*, *43*, 259-275.
- Townsend, J. T., Hu, G. G., & Evans, R. J. (1984). Modeling feature perception in brief displays with evidence for positive interdependencies. *Perception & Psychophysics*, *36*(1), 35-49.
- Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America*, *92*(2), 723-735.
- Wenger, M. J., & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1106-1118.
- Wickens, T. D. (1992). Maximum-likelihood estimation of a multivariate gaussian rating model with excluded data. *Journal of Mathematical Psychology*, *36*, 213-234.

Statistical power in the null-hypothesis tests and Bayesian model

Although a hierarchical Bayesian model has a number of desirable properties, if the statistics of interest do not enable reliable detection of underlying interactions, they are of (extremely) limited use. The statistical power of Δ , z , π , and χ^2 were evaluated by generating simulated identification-confusion data from known (Gaussian) GRT models exhibiting failures of perceptual independence, perceptual separability, and decisional separability

A.1. Perceptual Independence, Decisional Separability, and Report Independence

The relationship between χ^2 and π , on the one hand, and failure of perceptual independence or failure of decisional separability, on the other, were probed by simulating response frequency data from identification-confusion probabilities derived from a bivariate Gaussian ‘perceptual’ distribution and two linear decision bounds. Two sets of identification-confusion probabilities were calculated: one for values of ρ (correlation) ranging from -0.8 to 0.8 in steps of 0.1 , and for values of c_x (slope of the horizontal decision bound) ranging from -0.7 to 0.7 also in steps of 0.1 .¹ The ranges of parameter values probed are illustrated in Figure A.1.

For each set of probabilities, 100 sets of (multinomial random variable) response frequencies were generated. For each set of response frequencies, χ^2 and π test statistics were calculated. For each value of ρ and c_x , power was calculated as the proportion of test statistics that indicated a failure of report independence. Thus, because the tests were based on a 2×2 stimulus space (the ‘standard’ GRT protocol), the χ^2 tests had a single degree of freedom, so any χ^2 value greater than 6.64 ($p \leq 0.01$) was taken to indicate an underlying interaction. π statistics were calculated for each of the 3000 posterior samples of θ values taken (as described above), and values of π were calculated

¹ The non-zero slope values should be considered relative to the fact that the marginal x and y variances of the perceptual distribution were fixed at one.

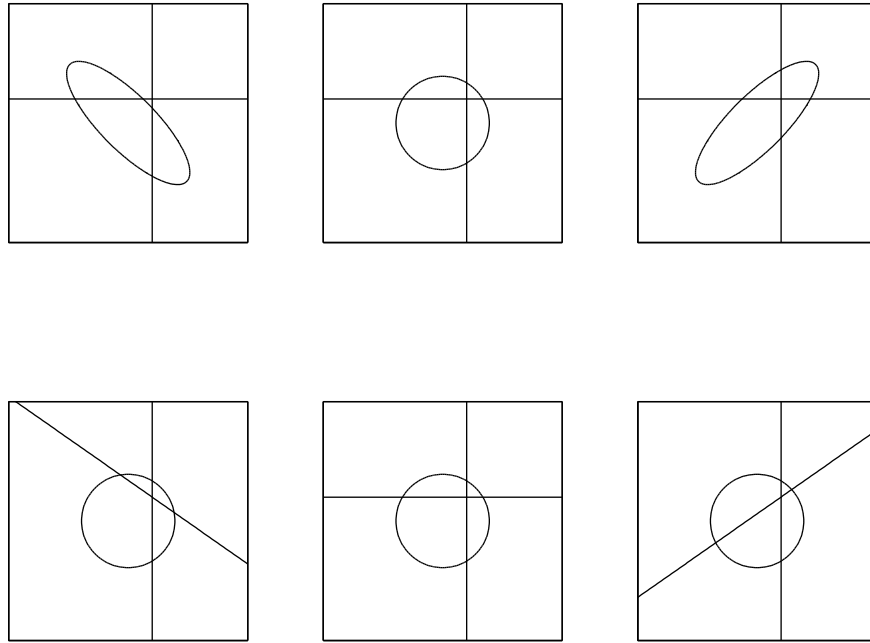


FIGURE A.1. Illustration of the range of parameter values employed in the simulation of data used to investigate the statistical power of χ^2 and π test statistics. The top row shows three values of the correlation coefficient ρ : -0.8 (left), 0 (middle), and 0.8 (right). The bottom row shows three values of the slope of the horizontal decision bound: -0.7 (left), 0 (middle), and 0.7 (right)

such that between the values, the most probable 99% of the posterior sample was contained (i.e., the marginal highest probability density region). If the region containing the most probable 99% of π values excluded zero on a given simulated data set, it was taken to indicate an underlying interaction.² Figure A.2 shows the results of this power analysis.

² Conservative criteria were employed here to ensure that the power analysis results are directly comparable to the results of data analyses presented below. Conservative criteria were employed in actual data analysis for two reasons: the data are numerous enough to support strong statistical inferences, and numerous tests (i.e., multiple comparisons) were carried out on each data set.

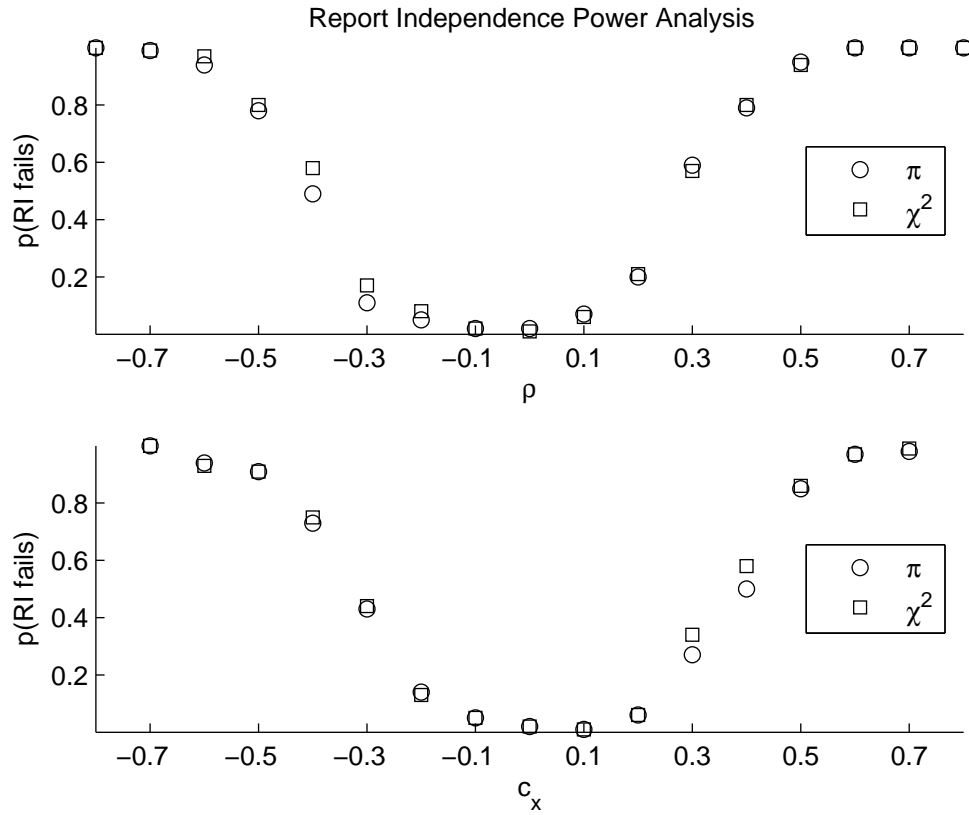


FIGURE A.2. Results of power analysis comparing the probability of detection of failure of report independence by χ^2 and π test statistics.

At all values of ρ and c_x , both methods detect underlying interactions equally well. There is a small divergence in the two methods in the small-to-middle range of negative ρ values and in the middle range of positive c_x values wherein π performs slightly worse than does χ^2 , which may be due either to idiosyncrasies of the simulated data or to a systematic inability of π to detect failures of report independence. In either case, it is beyond the scope of the present endeavor to determine the cause of these (small) divergences. This small difference must be weighed against the added benefits of a hierarchical Bayesian analysis as well as the additional information preserved in the signed statistic π that is not preserved by χ^2 .

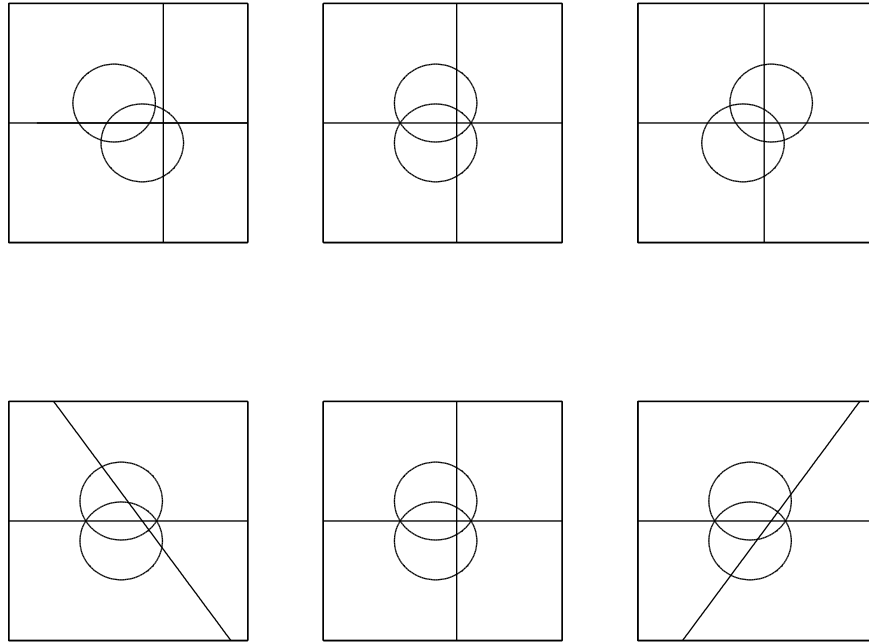


FIGURE A.3. Illustration of the range of parameter values employed in the simulation of data used to investigate the statistical power of z and Δ test statistics. The top row shows three values of μ_x : -1.4 (left), 0 (middle), and 0.2 (right). The bottom row shows three values of the slope of the vertical decision bound: -0.7 (left), 0 (middle), and 0.7 (right)

A.2. Perceptual Separability, Decisional Separability, and Marginal Response Invariance

The relationship between z and Δ , on the one hand, and failure of perceptual separability or failure of decisional separability, on the other, were probed by simulating response frequency data from identification-confusion probabilities derived from two (vertically aligned) bivariate Gaussian ‘perceptual’ distributions and one linear decision bound. Two sets of identification-confusion probabilities were calculated: one for values of μ_x of the ‘level two’ distribution ranging from -1.4 to 0.2 in steps of 0.1 (μ_x of the ‘level one’ distribution was fixed at -0.6), and for values of c_y (slope of the

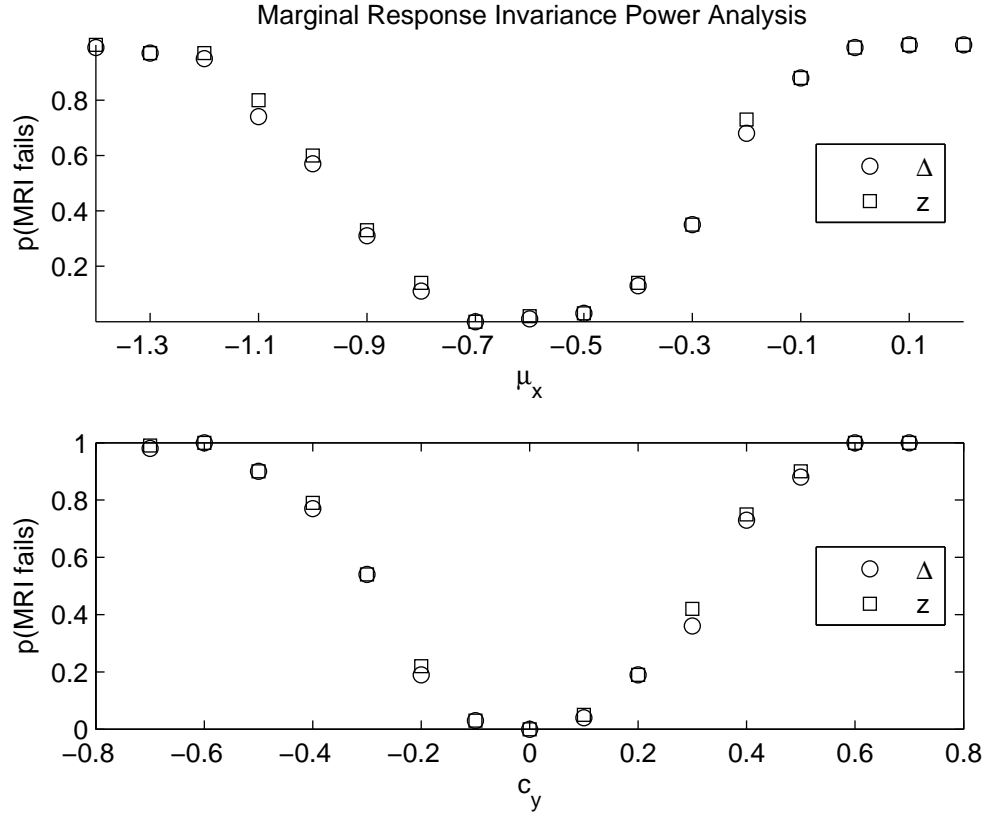


FIGURE A.4. Results of power analysis comparing the probability of detection of failure of marginal response invariance by z and Δ test statistics.

vertical decision bound) ranging from -0.7 to 0.7 also in steps of 0.1 . Figure A.3 shows the ranges of parameter values used in the simulations. All Δ and z statistics were calculated to test marginal response invariance at level one of the x dimension (i.e., to compare probabilities to the left of the vertical decision bound).

$$(11) \quad z_{a_1} = \frac{p(a_1|A_1, B_1) - p(a_1|A_1, B_2)}{\sqrt{p_*(1-p_*)\left(\frac{1}{n_{A_1, B_1}} + \frac{1}{n_{A_1, B_2}}\right)}}$$

The z statistic is defined in equation 11 for a test of marginal response invariance for a_1 responses. In the equation, n_{A_i, B_j} is the number of presentations of the A_i, B_j stimulus, and p_* is the proportion of a_1 responses for both stimuli (i.e., A_1, B_1 and A_1, B_2). The same statistic can be calculated for any level on any dimension with appropriate changes in subscripts.

For each set of probabilities, 100 sets of (multinomial random variable) response frequencies were generated. For each set of response frequencies, z and Δ test statistics were calculated. For each value of μ_x and c_y , power was calculated as the proportion of test statistics that indicated a failure of marginal response invariance. Any z value greater than 2.58 ($p \leq 0.01$) was taken to indicate an underlying interaction. Δ statistics were calculated for each of the 3000 posterior samples of θ values taken (as described above), and marginal highest probability density region for Δ was calculated. If this region excluded zero on a given simulated data set, it was taken to indicate an underlying interaction. Figure A.4 shows the results of this power analysis.

At all values of μ_x and c_y , both methods reliably detect the underlying interaction; there are no notable differences between Δ and z .

APPENDIX B

BRUGS Code

```
# THE MODEL.
modelstring = "
  model {
    # Likelihood function:
    for( sb in 1 : Nsubj ) {
      for( st in 1 : Nstim ) {
        data_table[ st + Nstim·(sb-1), 1: Nresp]
          ~ dmulti( theta[st + Nstim·(sb-1), 1:Nresp], Ntr[st + Nstim·(sb-1)] )
      }
    }
    # Prior:
    for( sb in 1 : Nsubj ) {
      for( st in 1 : Nstim ) {
        theta[st + Nstim·(sb-1), 1: Nresp] ~ ddirch( alpha[st,1: Nresp] )
      }
    }
    # hyperprior:
    for( st in 1 : Nstim ) {
      for( re in 1 : Nresp ) {
        alpha[st,re] dgamma( 3, .02)
      }
    }
  }
" # close quote for modelstring
```

```
# Write model to a file:
.temp = file("model.txt","w"); writeLines(modelstring,con=.temp); close(.temp)
# Load model file into BRugs and check its syntax:
modelCheck( "model.txt" )
modelData(fileName="data.txt")
nchain = 3
modelCompile( numChains = nchain )
modelGenInits()
# burn in
BurnInSteps = 10000 # try 20000 or so
modelUpdate( BurnInSteps )
# actual samples
samplesSet( c( "theta" , "alpha" ) )
stepsPerChain = 1100
thinStep = 200
modelUpdate( stepsPerChain , thin=thinStep )
```

APPENDIX C

Acoustic measurements of stimuli

Following Silbert and Jong (2008), the following spectral and temporal measurements were taken of the nonsense syllable stimuli. For the consonants, consonant duration, acoustic power below and above 750 Hz, the spectral mean, spectral variance, spectral skewness, and spectral kurtosis. For the vowels, F_0 at the consonant-vowel boundary, mid-vowel and in the portion of the vowel furthest in time from the consonant (i.e., the beginning for coda consonants, at the end for onset consonants), F_1 , and F_2 at consonant-vowel boundary and mid-vowel. The following figures show the results of these analyses. For the stops (i.e., [p], [b], [t], and [d]), consonants were defined, for onset consonants, as the portion between oral closure release and regular voicing onset, and for coda consonants the portion between offset of regular voicing and release of oral closure.

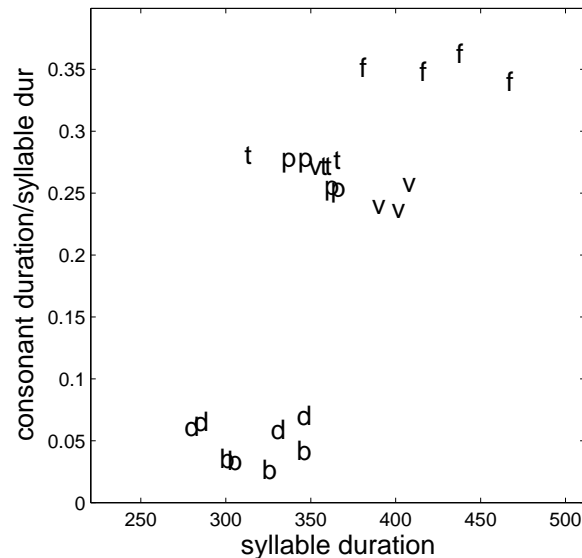


FIGURE C.1. Onset consonant duration (as a proportion of syllable duration) and syllable duration

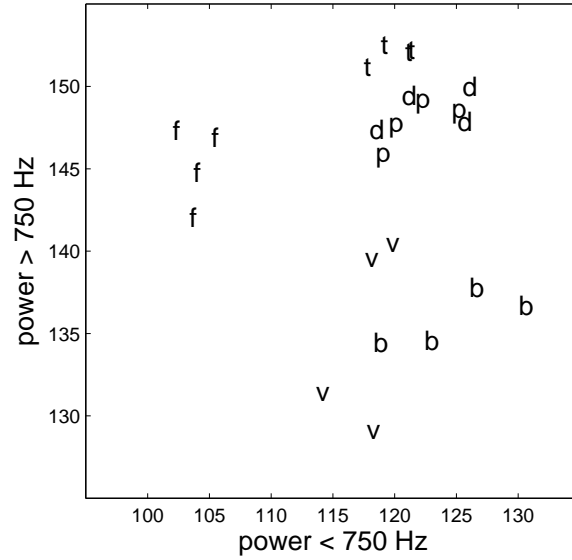


FIGURE C.2. Spectral power (in dB re: 1 bit) above (y -axis) and below (x -axis) 750 Hz for onset consonants

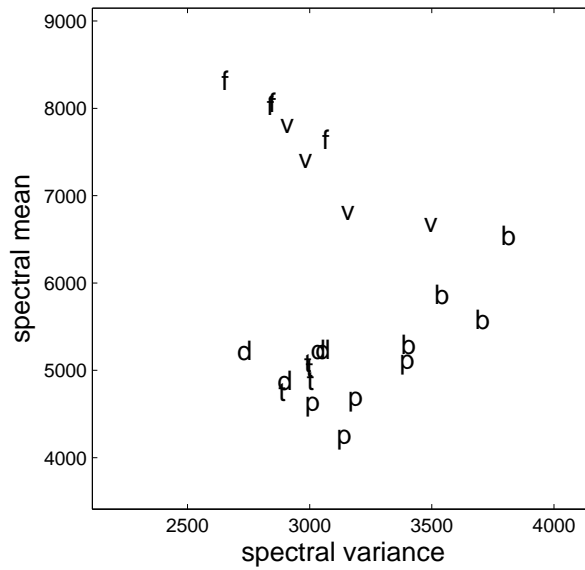


FIGURE C.3. Spectral means (y -axis) and variances (x -axis) for onset consonants

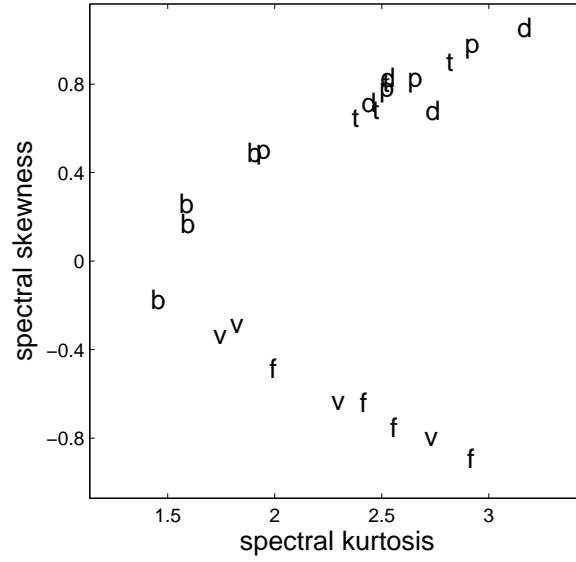


FIGURE C.4. Spectral skewness (y -axis) and spectral kurtosis (x -axis) for onset consonants

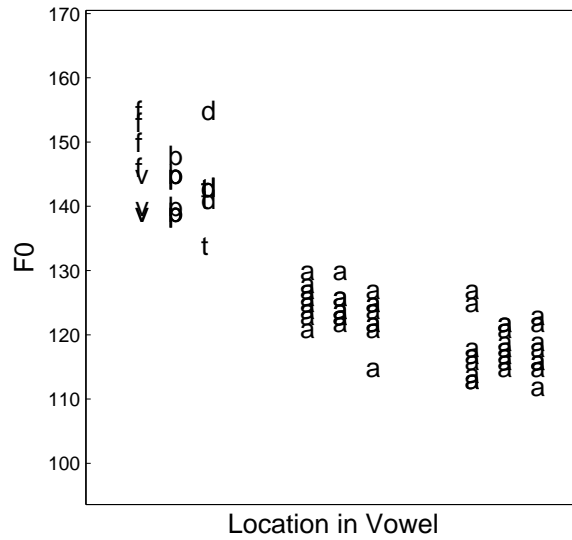


FIGURE C.5. Fundamental frequency (F0; y -axis) at voicing onset (consonant labels), mid-vowel, and at syllable end

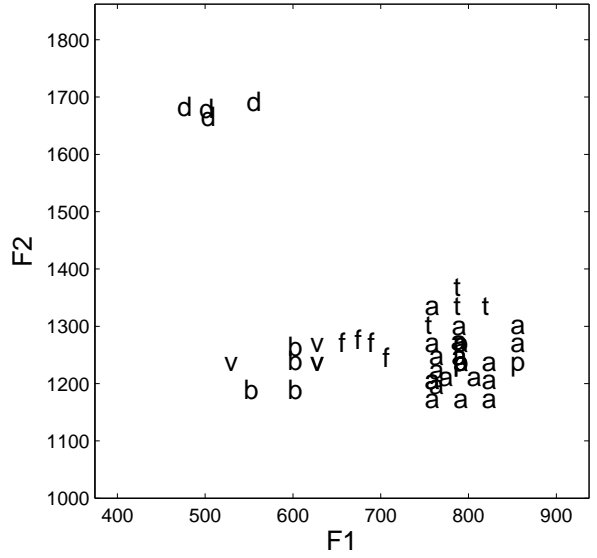


FIGURE C.6. First (F1; x -axis) and second (F2; y -axis) formants at voicing onset (consonant labels) and mid-vowel

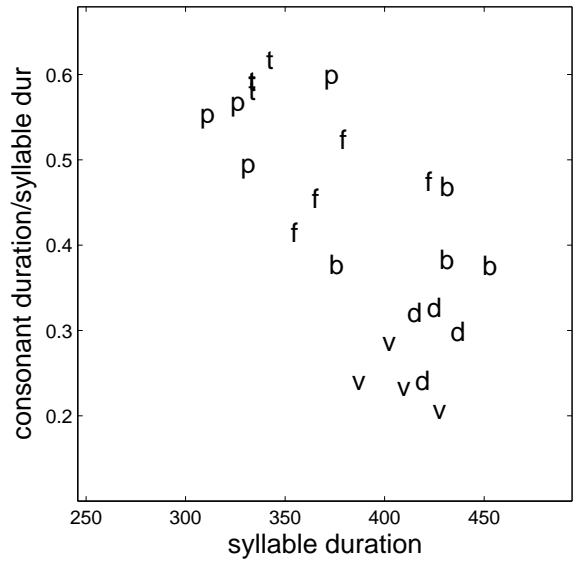


FIGURE C.7. Coda consonant duration (as a proportion of syllable duration) and syllable duration

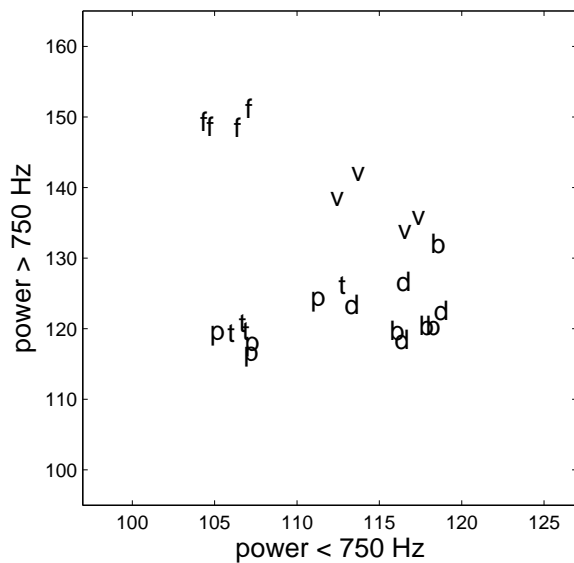


FIGURE C.8. Spectral power (in dB re: 1 bit) above (y -axis) and below (x -axis) 750 Hz for coda consonants

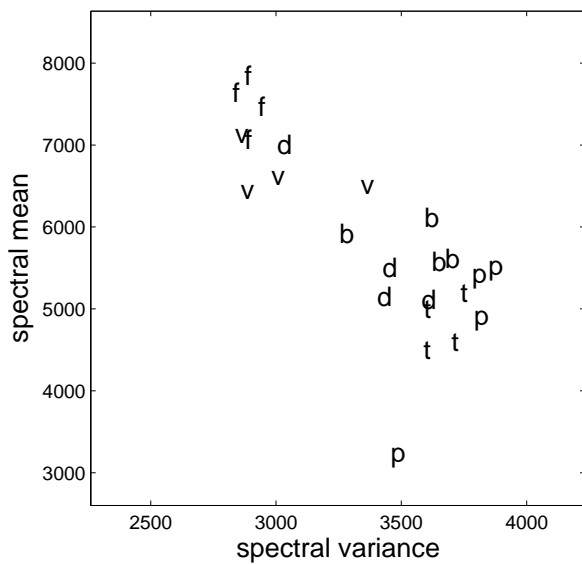


FIGURE C.9. Spectral means (y -axis) and variances (x -axis) for coda consonants

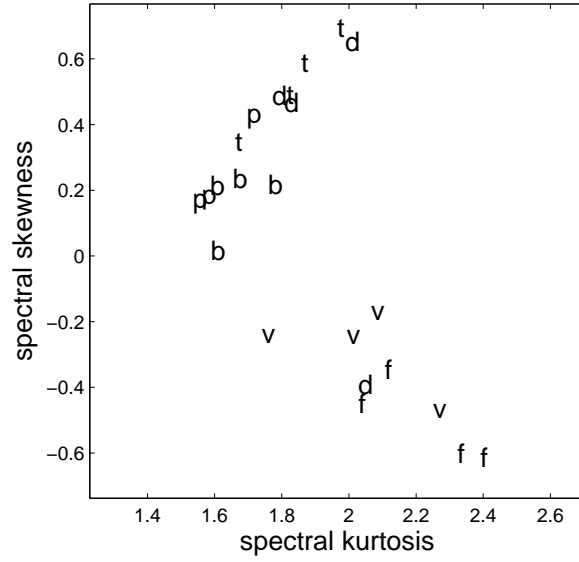


FIGURE C.10. Spectral skewness (y -axis) and spectral kurtosis (x -axis) for coda consonants

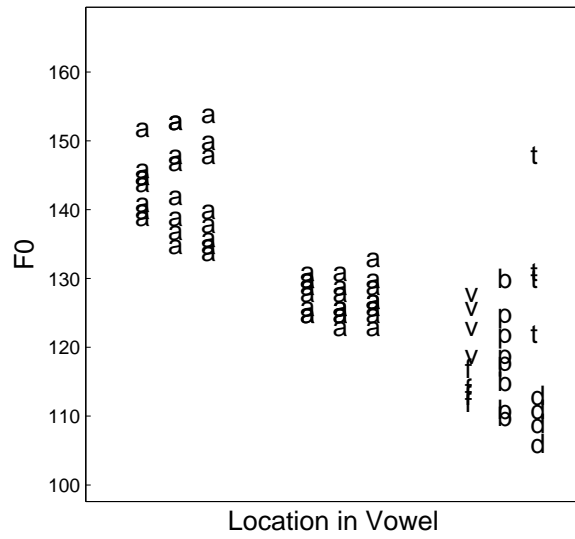


FIGURE C.11. Fundamental frequency (F0; y -axis) at voicing onset, mid-vowel, and voicing offset (consonant labels)

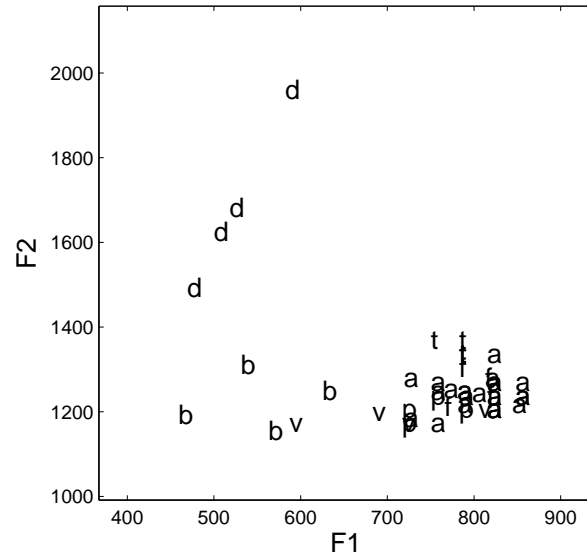


FIGURE C.12. First (F1; x -axis) and second (F2; y -axis) formants at voicing coda (consonant labels) and mid-vowel

Noah H. Silbert

Postdoctoral Fellow
Mathematical Psychology Laboratory
Department of Psychological & Brain Sciences
1101 E. 10th Street
Indiana University
Bloomington, IN 47405
(812) 855-1554
nosilber at indiana dot edu
www.nhsilbert.net

Education

B.A. with distinction in Linguistics, Indiana University, 1998
M.A. in English as a Second Language, University of Hawaii, 2002
Ph.D. in Linguistics and Cognitive Science, Indiana University, 2009
M.S. in Applied Statistics, Indiana University, 2010

Linguistics Coursework

Introduction to Historical Linguistics
Introductory Phonetics
Phonological Analysis
Syntactic Analysis
Advanced Phonetics
Advanced Phonological Description
Advanced Readings in Linguistics: Second Language Phonetics and Phonology
Seminar in Phonetics: Individual Differences

Cognitive Science Coursework

Programming Methods in Cognitive Science
Philosophical Foundations of Cognitive Science
The Brain and Cognition
Experimental Methods in Cognitive Science

Mathematical Modeling and Statistics Coursework

Applied Linear Models I, II
Bayesian Theory and Data Analysis
Multivariate Data Analysis
Introduction to Statistical Theory
Digital Signal Processing
Models in Cognitive Science
Introduction to Dynamical Systems in Cognitive Science
Advanced Statistics in Psychology I, II
Representation of Structure in Psychological Data
Introduction to Mathematical Psychology

Research Interests

Speech perception, psychoacoustics, phonological structure in speech production and perception, audio-visual information integration, auditory categorization, second language speech perception, models of perception and decision making

Research Experience

October, 2009 – Current: Postdoctoral Research Fellow, NIH Grant, Perceptual Processes, Evidence, and Configurality, Principal Investigator James T. Townsend, jtownsen @ indiana.edu

September, 2007 – September, 2009: Graduate Research Assistant, NIH Grant, Perceptual Processes, Evidence, and Configurality, Principal Investigator James T. Townsend, jtownsen @ indiana.edu

September, 2006 – August, 2007: NIH Pre-doctoral Fellow, Cognitive Modeling Training Grant, director James T. Townsend, jtownsen @ indiana.edu

May, 2006 – August, 2006: Graduate Research Assistant, Audiology Research Laboratory, director Larry Humes, humes @ indiana.edu

January, 2006 – May, 2006: Graduate Research Assistant, NSF Project, Prosody in Cross-Language Production and Perception, Principal Investigator Kenneth de Jong, kdejong @ indiana.edu

September, 2003 – December, 2005: Graduate Research Assistant, Linguistic Speech Laboratory, director Kenneth de Jong, kdejong @ indiana.edu

Teaching Experience

September, 2007 – May, 2008; September, 2008 – May, 2009: Grader, Introduction to Bayesian Data Analysis I & II, Department of Psychological and Brain Sciences, taught by Professor John Kruschke, kruschke @ indiana.edu

January, 2007 – May, 2007; September, 2008 – December, 2008: Assistant Instructor, Introduction to Mathematical Psychology, Department of Psychological and Brain Sciences, taught by Professor James T. Townsend, jtownsen @ indiana.edu

September, 2005 – December, 2005: Instructor, Introduction to the Study of Language, Department of Linguistics, Indiana University.

July, 2005 – August, 2005: Instructor, Statistical Techniques, Department of Psychological and Brain Sciences, Indiana University.

January, 2005 – May, 2005: Assistant Instructor, Introductory Phonetics, taught by Professor Robert Port, port @ indiana.edu

September, 2004 – December, 2004: Instructor, Phonological Analysis, Department of Linguistics, Indiana University.

September, 2003 – December, 2003; January, 2004 – May, 2004: Assistant Instructor, Introduction to the Study of Language, taught by Professor Samuel Obeng (2003), sobeng @ indiana.edu, taught by Sean McLennan (2004).

July, 2001 - August, 2001: Nihon University, Mishima, Japan, and Outreach College. Teaching Assistant, pronunciation and oral fluency, and Student Interchanger, ESL conversation tutoring. Supervisor Judy Ensing, (808) 956-3414

Publications

Peer Reviewed

Silbert, N. H., de Jong, K., Lentz, J. J., & Townsend, J. T. (*in preparation*) Syllable structure and integration of voicing and place of articulation information in stop-consonant identification.

Silbert, N. H., de Jong, K., Lentz, J. J., & Townsend, J. T. (*in preparation*) Syllable structure and Integration of voicing and manner of articulation information in labial consonant identification.

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

Silbert, N. H., de Jong, K., 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, 339-343.

de Jong, K., **Silbert, N.**, & Park, H. (2009). Generalization across segments in second language consonant identification. *Language Learning*, 59 (1), 1-31.

Silbert, N., & de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America*, 123 (5), 2769-2779.

Silbert, N., & de Jong, K. (2007). Laryngeal Feature Structure in 1st and 2nd Language Speech Perception. In J. Trouvain and W.J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarland University, Saarbrücken, Germany. pp.1901-1905. Available at www.icphs2007.de/conference/book_of_abstracts.pdf

Conference Presentations

Silbert, N. H., de Jong, K. J., Lentz, J. J., & Townsend, J. T. (2009). Integration of phonological information in obstruent consonant identification. Poster presented at the *158th Meeting of the Acoustical Society of America*, San Antonio, Texas.

Vandam, M., **Silbert, N. H.**, & Port, R. F. (2009). Fine-grained control of voice-onset time production, lexical usage-frequency, and phrasal context. Poster presented at the *158th Meeting of the Acoustical Society of America*, San Antonio, Texas.

- Townsend, J. T., Houpt, J. W., **Silbert, N. H.**, & Altieri, N. A. (2009). Toward a new psychophysics based on geometry and stochastic systems theory. Keynote address presented at and appearing in the proceedings of the *25th Annual Meeting of the International Society for Psychophysics*, Galway, Ireland.
- Altieri, N., **Silbert, N.**, & Townsend, J. T. (2009). An analysis of perceptual dependencies in audiovisual speech perception: A new approach. Talk presented at and appearing in the proceedings of the *25th Annual Meeting of the International Society for Psychophysics*, Galway, Ireland.
- Altieri, N., **Silbert, N.**, & Pei, L. (2009). An investigation of perceptual dependencies in audiovisual speech perception. Poster presented at *10th International Multisensory Research Forum*, New York.
- Silbert, N.**, Townsend, J. T., & Lentz, J. J. (2008). Three types of decision bounds in General Recognition Theory. Talk given at the *41st Annual Society for Mathematical Psychology Conference*, Washington, D.C.
- de Jong, K., Hao, Y.-C., Park, H., & **Silbert, N.** (2008). The ‘[+English]’ effect in acquisition of English as a second language: Novel- category biases in second language perception and production. *Journal of the Acoustical Society of America*, 123, 3332. (Poster Abstract)
- Silbert, N.**, & de Jong, K. (2007). Response bias, type and token frequency, and prosodic context in segment identification. *Journal of the Acoustical Society of America*, 122, 3019. (Abstract).
- Silbert, N.**, Townsend, J. T., & Lentz, J. J. (2007). Independence in the perception of complex non-speech sounds. *Journal of the Acoustical Society of America*, 122, 3062. (Abstract).
- Silbert, N.**, & de Jong, K. (2007). Distinctive features and mathematical models of perception and decision making. Poster presented at *Where Do Features Come From? Phonological Primitives in the Brain, the Mouth, and the Ear*, University of Paris III, Paris, France.
- Silbert, N.**, & de Jong, K. (2007). Laryngeal Feature Structure in 1st and 2nd Language Speech Perception. Talk given at the *16th International Congress of Phonetic Sciences*, Saarland University, Saarbrücken, Germany.
- Silbert, N.**, Townsend, J. T., & Lentz, J. J. (2007). Independence in the perception of complex non-speech sounds. Poster presented at the *40th Annual Meeting of the Society for Mathematical Psychology*, University of California, Irvine.
- de Jong, K.J., & **Silbert, N.** (2005). On Segmental Factorability in Second Language Learning. Paper B10 presented at the *First ASA Workshop on Second Language Acquisition*, 15 May, 2005, Vancouver, B.C., Canada.

de Jong, K., **Silbert, N.**, & Park, H. (2004). Segments and segmental properties in cross-language perception: Korean perception of English obstruents in various prosodic locations. *Journal of the Acoustical Society of America*, 115, 2504. (Abstract).

Park, H., de Jong, K., & **Silbert, N.** (2004). Cross-language perceptual category mapping: Korean perception of English obstruents. *Journal of the Acoustical Society of America*, 115, 2504. (Abstract).

Technical Reports

de Jong, K.J., & **Silbert, N.** (2006). On Segmental Factorability in Second Language Learning. *IULC Working Papers – Online*, 6-02.

Silbert, N., de Jong, K., & Park, H. (2005). Linguistic Generalization in L2 Consonant Identification Accuracy: a Preliminary Report. *IULC Working Papers – Online*, 5-01.

Professional Service

Reviewer, Language Learning, Editor Robert DeKeyser

Reviewer, Journal of Phonetics, Editor Stephan A. Frisch

Reviewer, Speech Communication, Guest Editor Maria Luisa Garcia Lecumberri

Professional Memberships

Acoustical Society of America

American Statistical Association

Cognitive Science Society

Linguistic Society of America

Society for Mathematical Psychology