Chapter 13

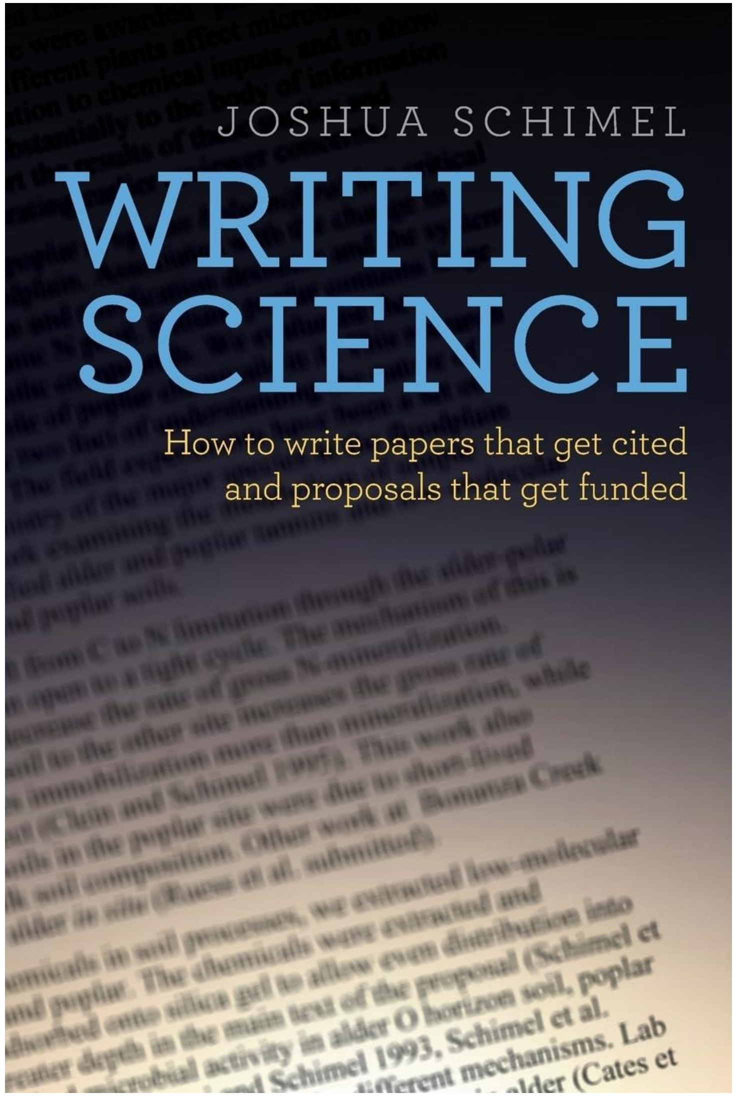
Lin 205C
Santiago Barreda

Writing a Paper

- Writing a paper is like directing a movie.
- A movie is made up of hundreds or thousands of discrete ‘shots’, stuck together to give the impression of a single, continuous, coherent story.
- Your paper is a bunch of random facts, numbers, and statements.
- You need to put this together to form a coherent and compelling story.

Writing a Paper

- Think of writing as a technical skill like scripting
- You need people to read (and understand) your papers
- Papers don't have to be boring and rigid



Writing a Paper

Opening (O): Whom is the story about? Who are the characters? Where does it take place? What do you need to understand about the situation to follow the story? What is the larger problem you are addressing?

Challenge (C): What do your characters need to accomplish? What specific question do you propose to answer?

Action (A): What happens to address the challenge? In a paper, this describes the work you did; in a proposal, it describes the work you hope to do.

Resolution (R): How have the characters and their world changed as a result of the action? This is your conclusion—what did you learn from your work?

Writing a Paper

LD Structure: ABDCE front-loads the story more than OCAR by collapsing the challenge into the opening, but some audiences are so impatient they won't stick around for a resolution. For them, you need to intensify the front-loading. The most extreme case of this is used by newspaper reporters. Reporters use a structure that they call the "inverted pyramid"; I call it Lead/Development (LD) to highlight its key functional elements. In LD structure, the core of the story is in the first sentences (the lead, L) and the rest fills out and develops the story (the development, D). In LD structure, the lead collapses the opening, challenge, *and* resolution into a single short section, possibly as little as a single sentence.

OCAR:	*Slowest—take your time working into the story.*
ABDCE:	*Faster—get right into the action.*
LDR:	*Faster yet—but people will read to the end.*
LD:	*Fastest—the whole story is up front.*

First Sentences

- “Current public health guidelines in the United States, the United Kingdom, and Australia recommend that women consume a supplemental dose of 400 μg of folic acid per day in the month preceding and during the first trimester of pregnancy to reduce the risk of neural tube defects in children.”

- “In meiosis, genes that are always transmitted together are described as showing “linkage.” Linkage, however, can be incomplete, due to the exchange of segments of DNA when chromosomes are paired. This incomplete linkage can lead to the creation of new pairings of alleles, creating new lineages with distinct sets of traits.”

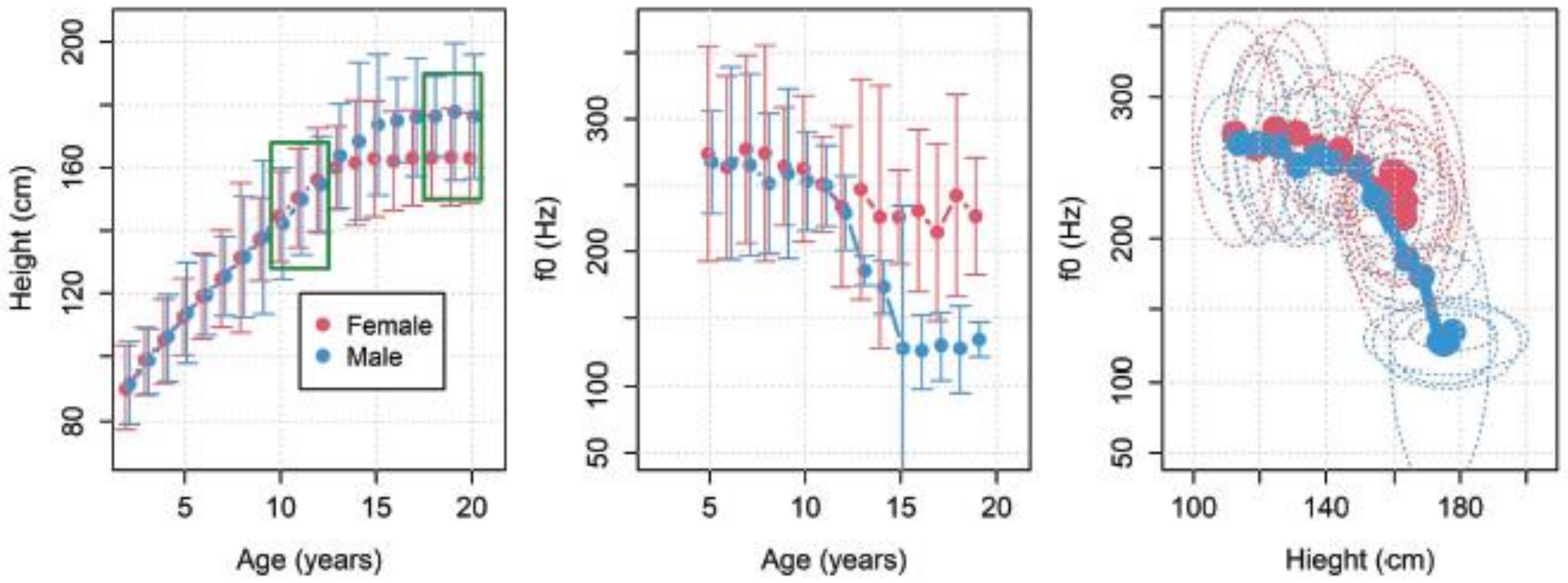
Perception of Speaker Characteristics

- Understand perception of apparent:

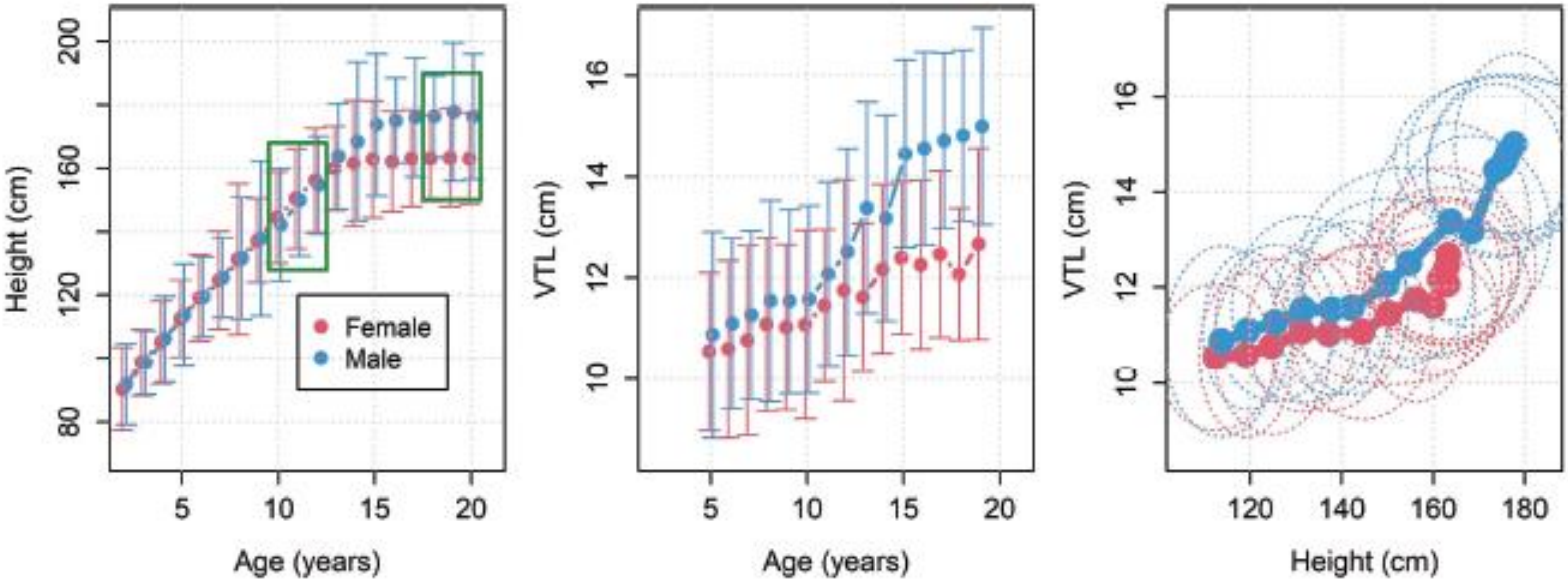
  - Gender (male or female)
  - Age (child or adult)
  - Height (in feet and inches)

- Based on:
  - pitch (f0)
  - vocal tract-length (formants)

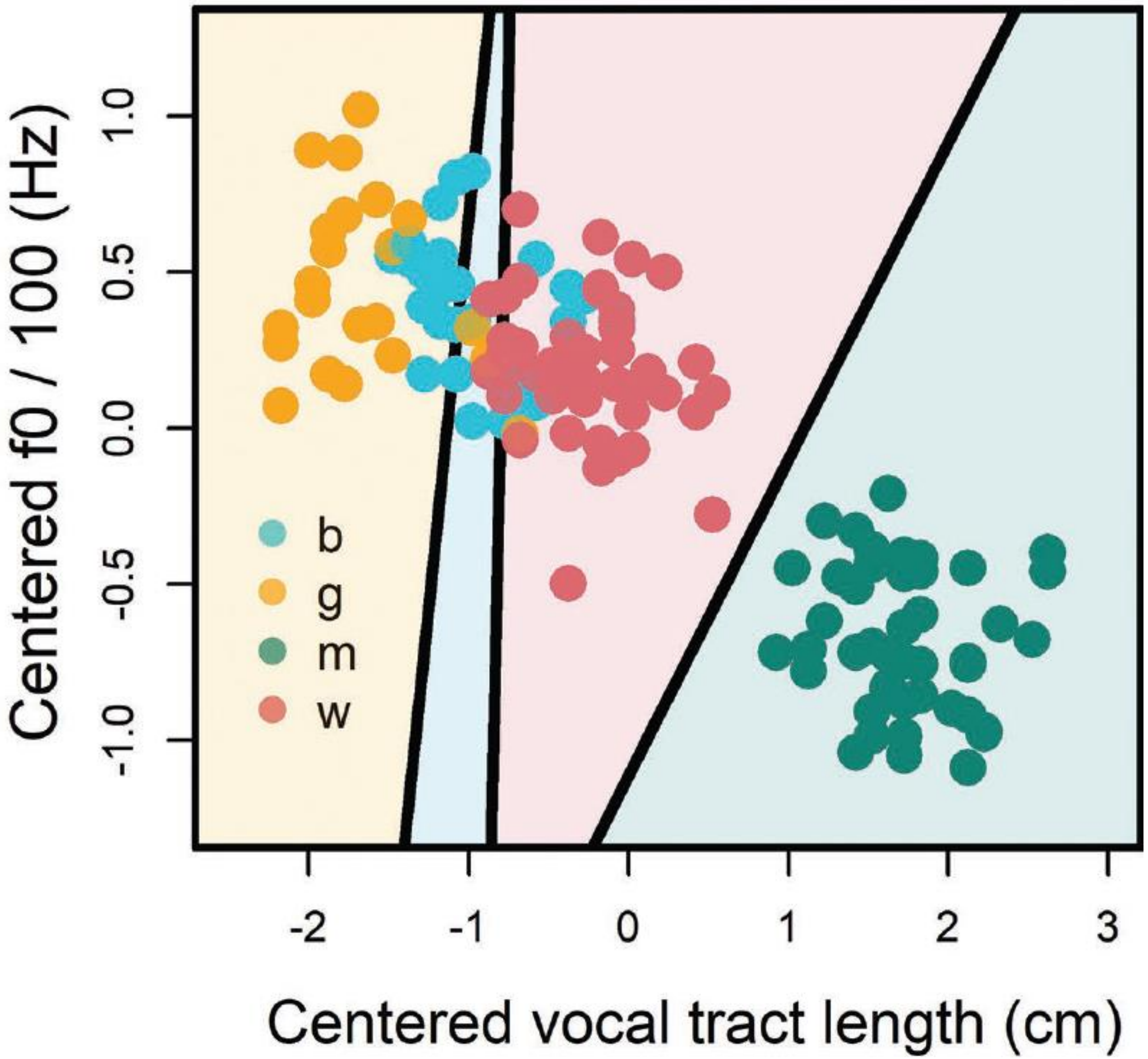
Perception of Speaker Characteristics



Perception of Speaker Characteristics



Perception of Speaker Characteristics



Perception of Speaker Characteristics

(Q1) What is the relationship between speech acoustics, specifically speaker f0 and VTL, and apparent height/gender?

(Q2) Do the effects for speaker f0 and VTL on apparent height/gender vary based on apparent speaker characteristics?

```
characteristic ~ f0 + vtl
```

Size Perception

Contents lists available at ScienceDirect

Journal of Phonetics

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

ELSEVIER

Inaccurate but predictable: Vocal-tract length estimation and gender stereotypes in height perception

Check for updates

Santiago Barreda [a,*], Kristin Predeck [a,b]

[a] *Department of Linguistics, University of California, Davis, CA 95616, USA*
[b] *Amazon AGI – Information, United States*

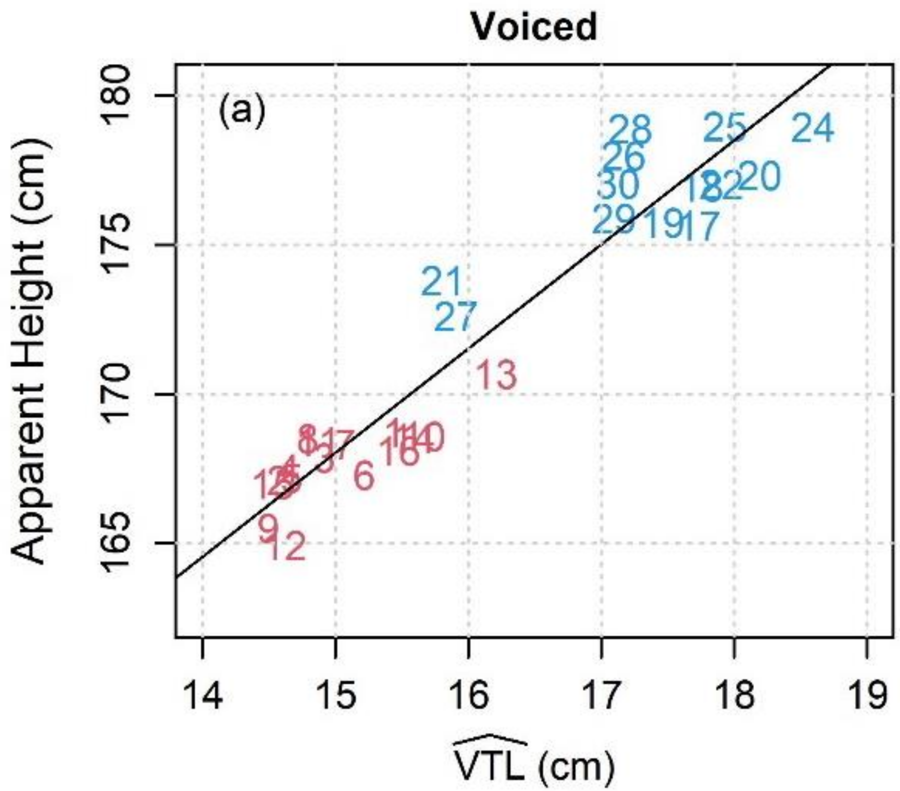
A R T I C L E   I N F O

Article history:
Received 15 June 2023
Received in revised form 31 October 2023
Accepted 8 December 2023
Available online 2 January 2024

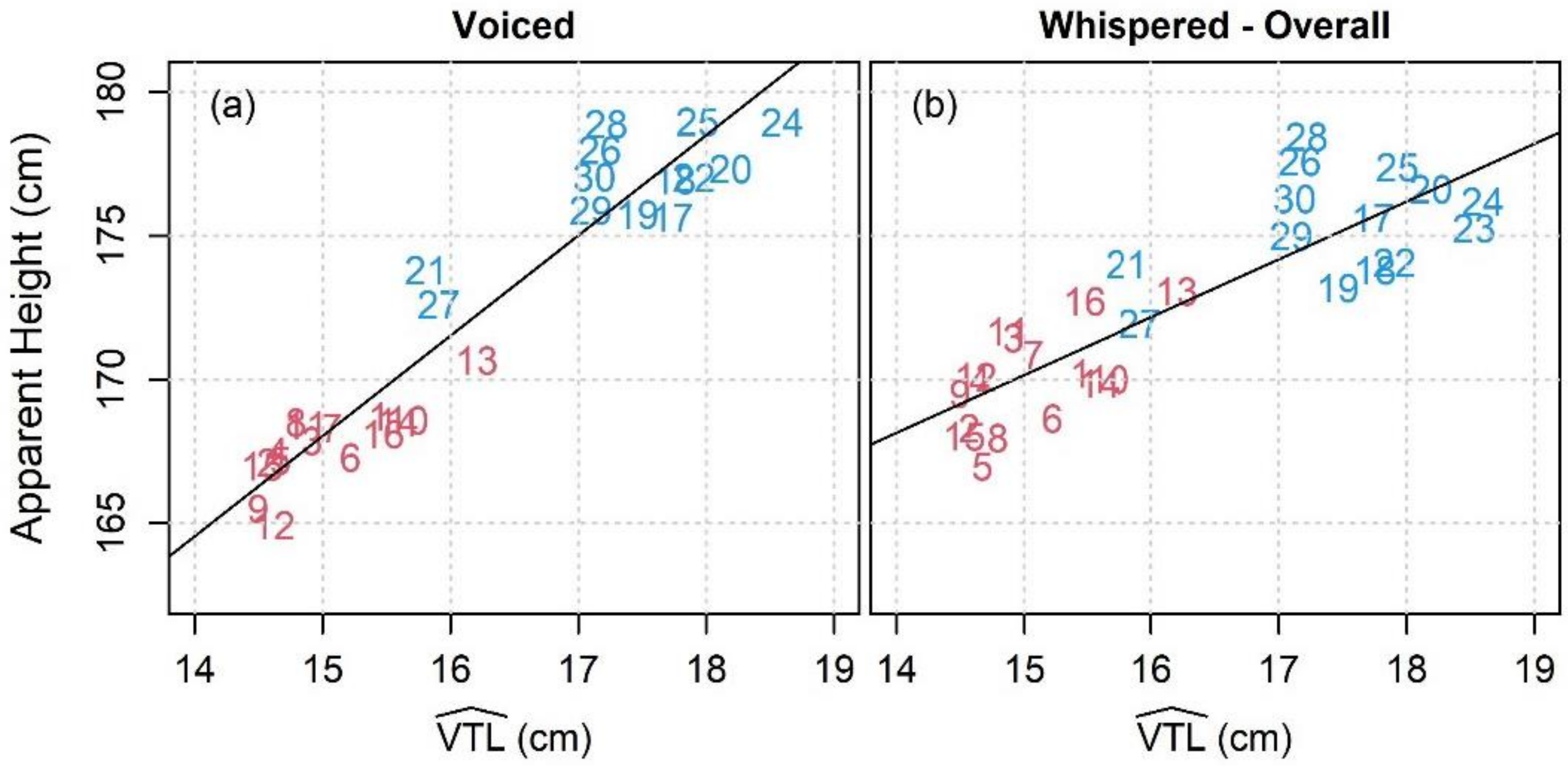
A B S T R A C T

Research suggests human listeners are not very accurate in assessing the size of adults from their speech, though they appear to be consistent in their judgments across listeners. Two experiments were conducted to investigate the importance of the higher formants for providing consistent height judgments, how consistent these height judgments are across replications, and the role of f0 and social knowledge in maintaining the stability of apparent speaker height judgments. Listeners were presented with syllables produced by 30 adult male and female speak-

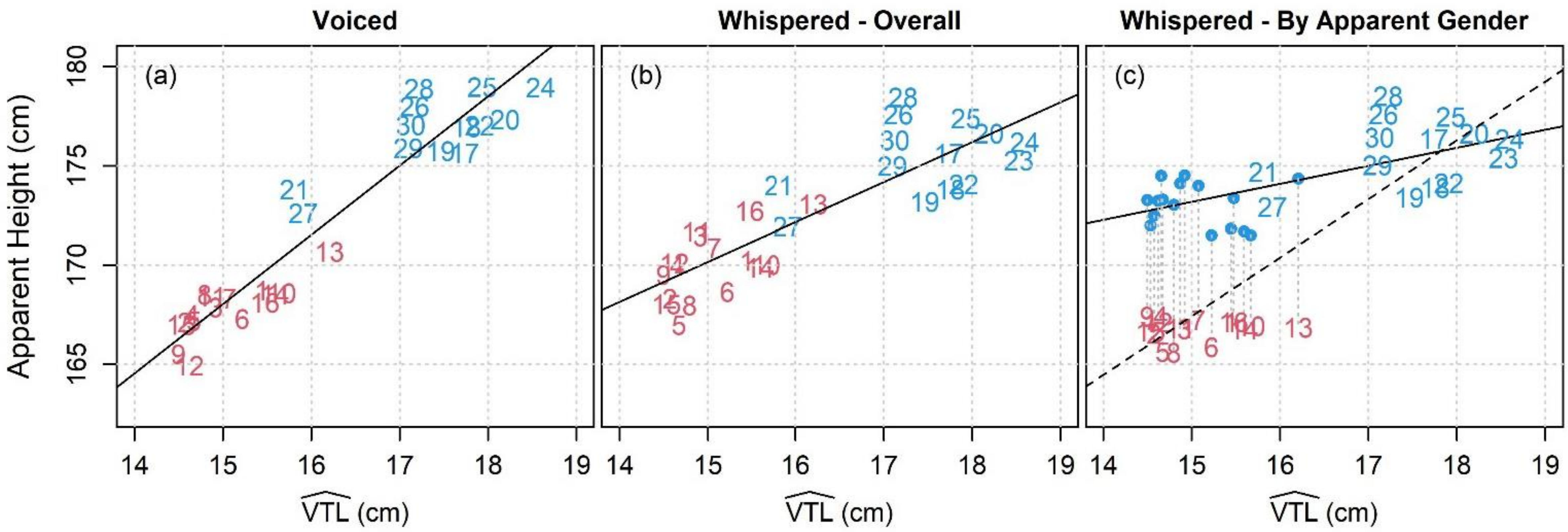
Size Perception



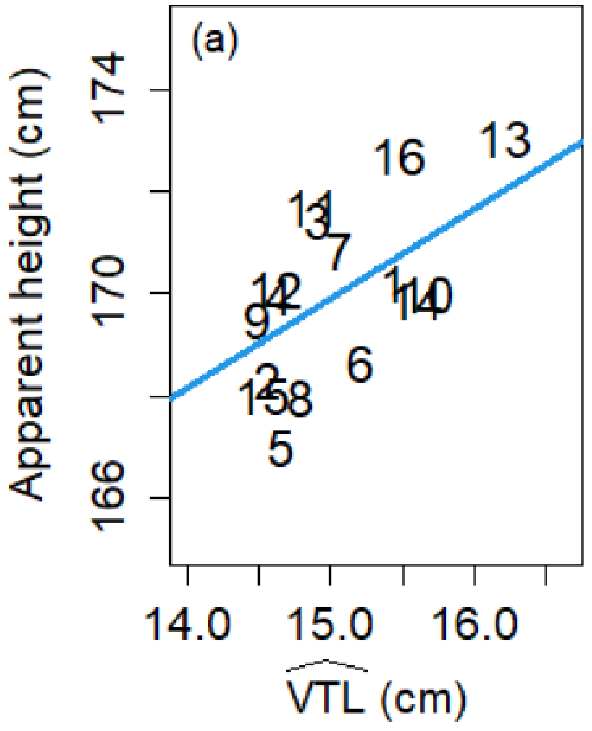
Size Perception



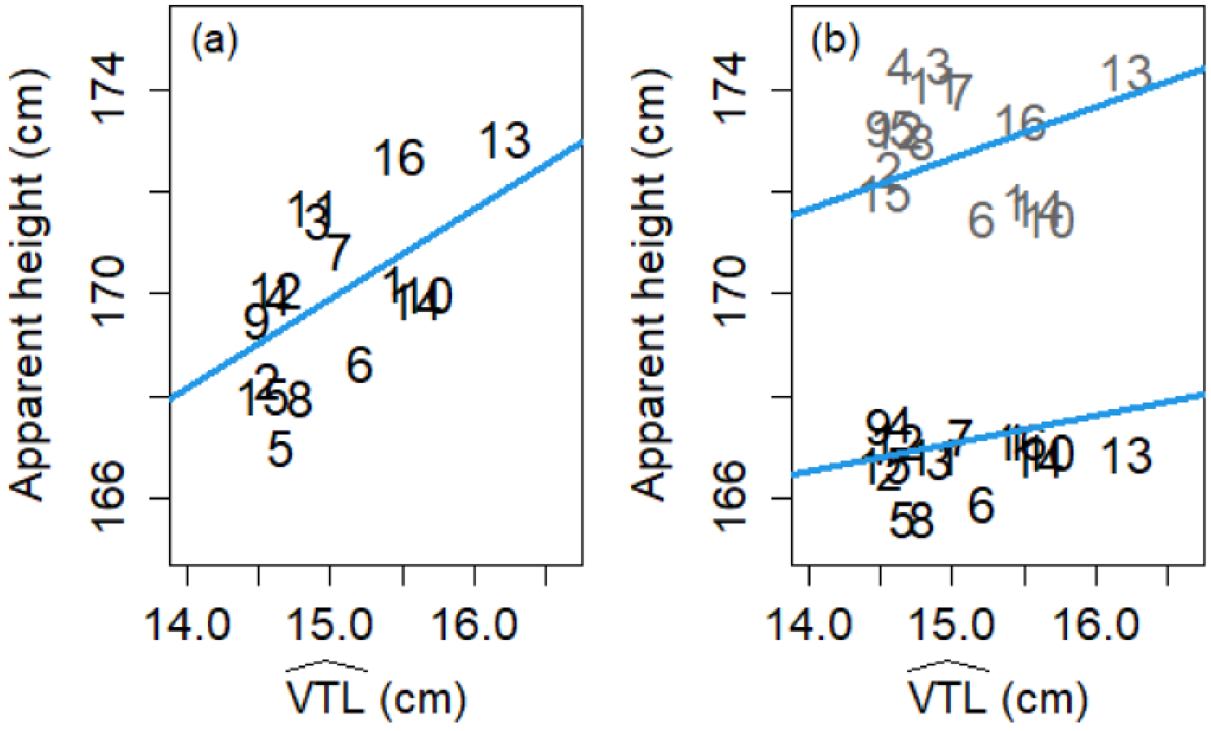
Size Perception



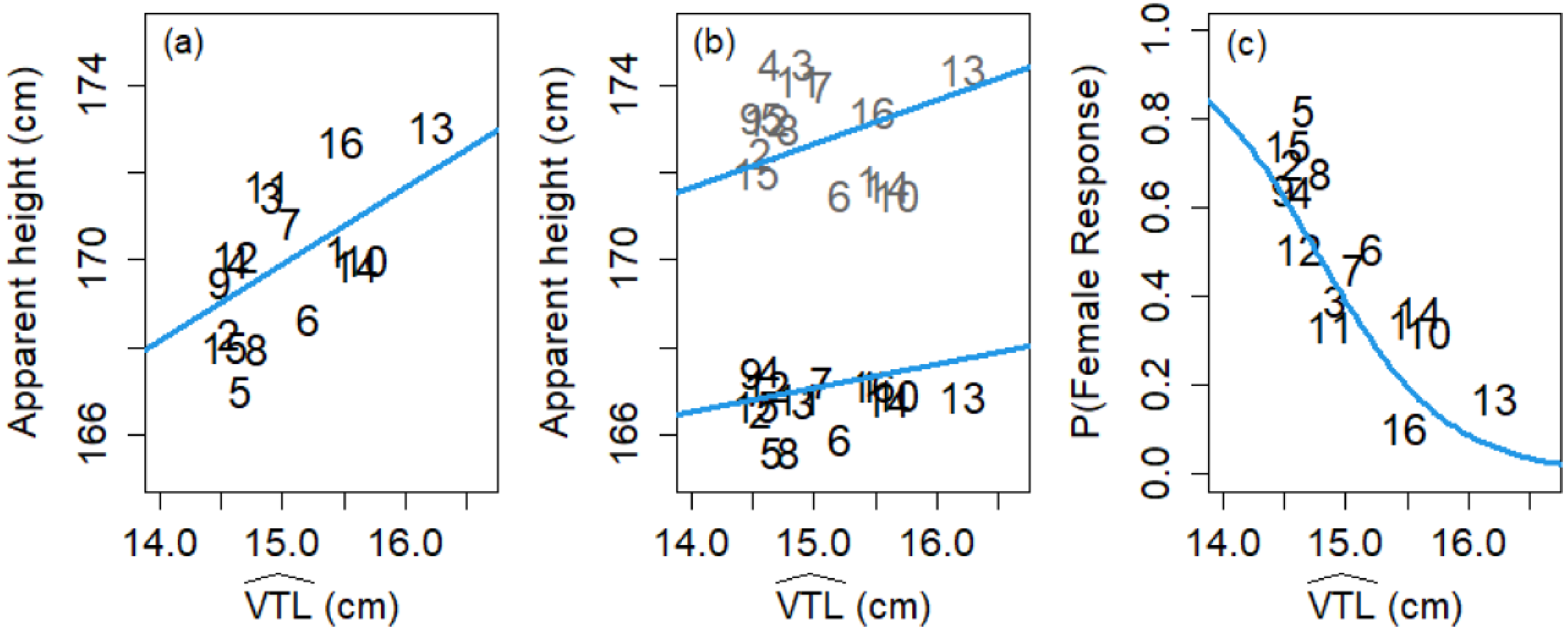
Size Perception



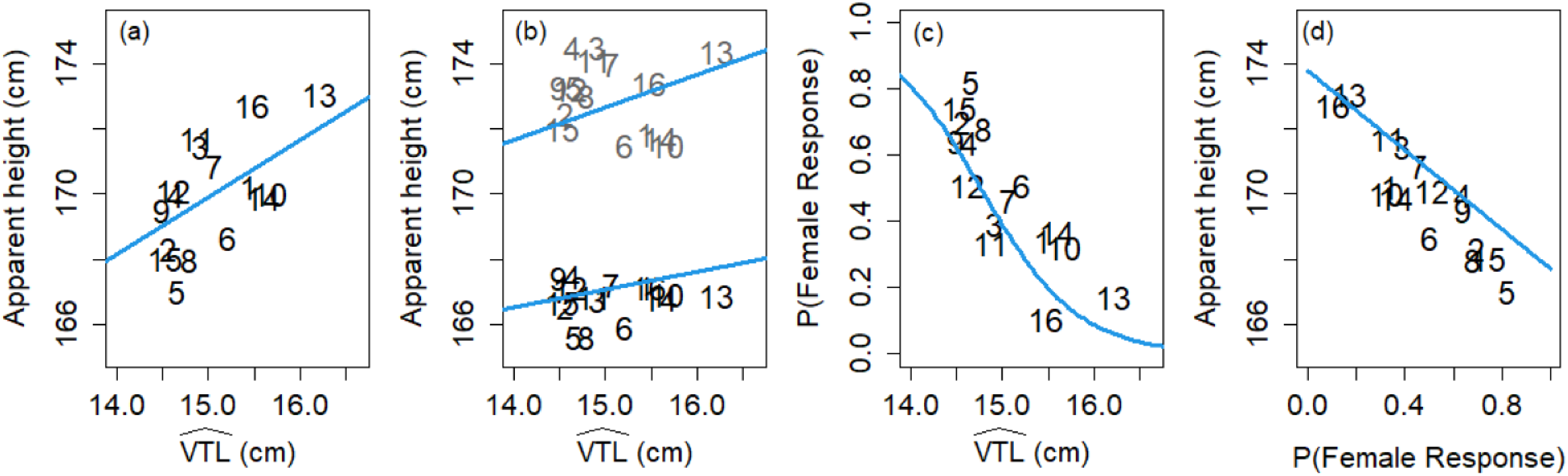
Size Perception



Size Perception



Size Perception



Gender Perception

JASA ARTICLE

Perception of gender in children's voices

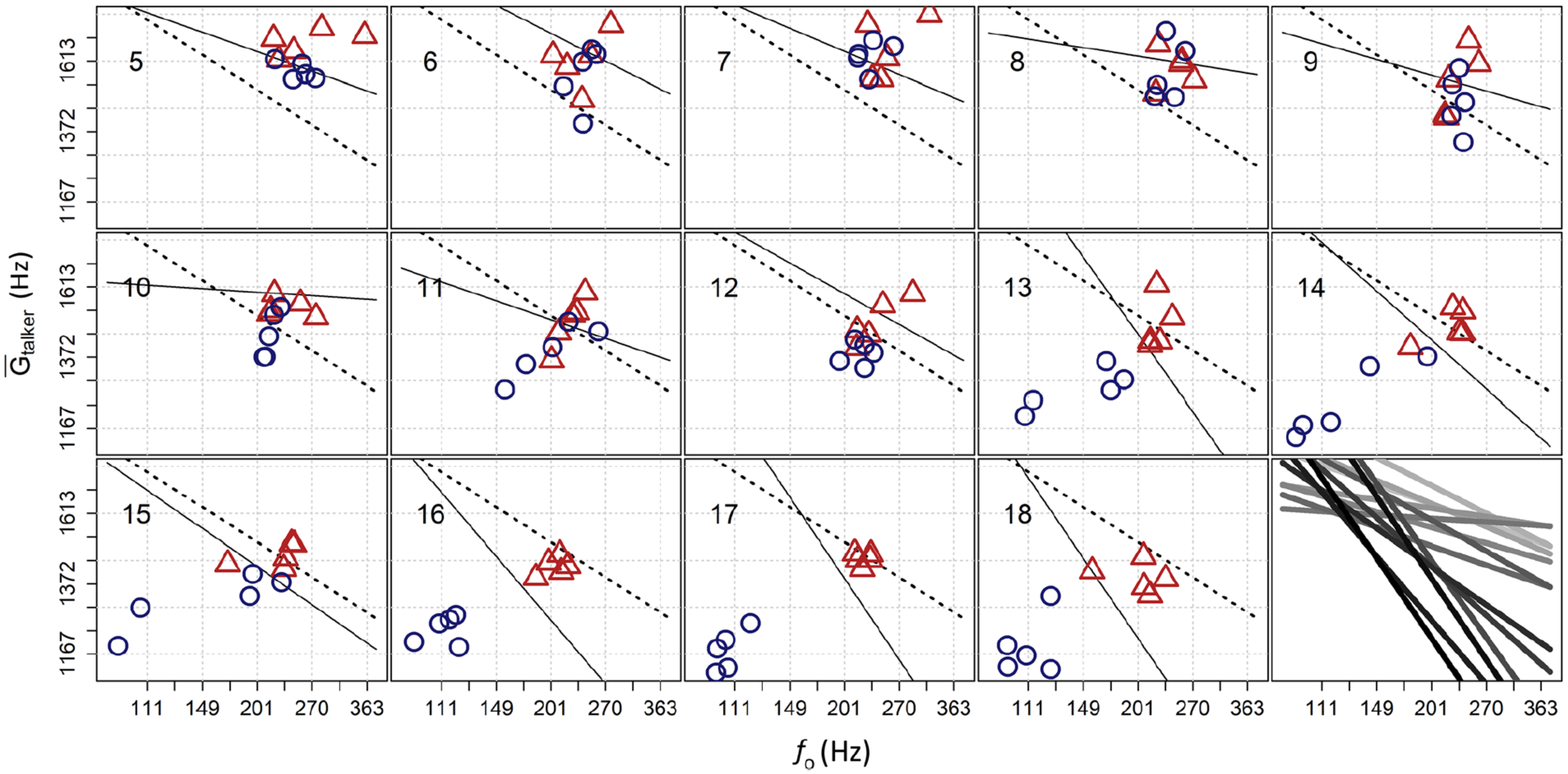
Santiago Barreda[1,a)] and Peter F. Assmann[2]
[1]*Department of Linguistics, University of California, Davis, California 95616, USA*
[2]*School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, Texas 75080, USA*

ABSTRACT:
To investigate the perception of gender from children's voices, adult listeners were presented with /hVd/ syllables, in isolation and in sentence context, produced by children between 5 and 18 years. Half the listeners were informed of the age of the talker during trials, while the other half were not. Correct gender identifications increased with talker age; however, performance was above chance even for age groups where the cues most often associated with gender differentiation (i.e., average fundamental frequency and formant frequencies) were not consistently different between boys and girls. The results of acoustic models suggest that cues were used in an age-dependent manner, whether listeners were explicitly told the age of the talker or not. Overall, results are consistent with the hypothesis that talker age and gender are estimated jointly in the process of speech perception. Furthermore, results show that the gender of individual talkers can be identified accurately well before reliable anatomical differences arise in the vocal tracts of females and males. In general, results support the notion that the transmission of gender information from voice depends substantially on gender-dependent patterns of articulation, rather than following deterministically from anatomical differences between male and female talkers.
https://doi.org/10.1121/10.0006785

(Received 30 December 2020; revised 28 September 2021; accepted 30 September 2021; published online 23 November 2021)

[Editor: Jody Kreiman] Pages: 3949–3963

Gender Perception



Perception of Speaker Characteristics

(Q1) What is the relationship between speech acoustics, specifically speaker f0 and VTL, and apparent height/gender?
(Q2) Do the effects for speaker f0 and VTL on apparent height/gender vary based on apparent speaker characteristics?

```
characteristic ~ f0 + vtl
```

```
characteristic ~ (f0 + vtl + X) * (Y * Z)
```

Introduction

- Set up the:
  - Problem
  - Theory
  - Perspective
- Engage your reader!

Methods

- What did you do?
  - Participants
  - Stimuli
  - Procedure
  - Data Analysis (sometimes)
  - Everything else you did

Apparent Height and Gender

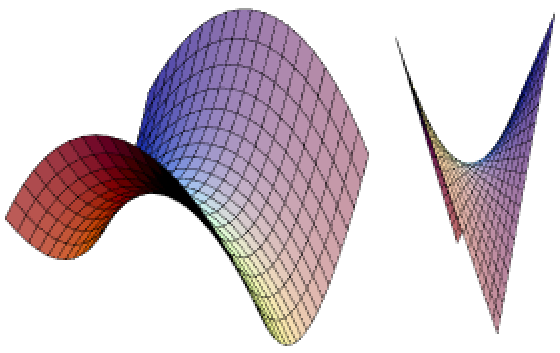
- Our Models:

```
Female ~ (vtl + f0) * A + ((vtl + f0) * A | L) + (1 | S)
```

```
height ~ (vtl+f0) * A * G + ((vtl+f0) * A * G | L) + (1 | S)
```

Hyperbolic Paraboloid

Download
Wolfram Notebook



A hyperbolic paraboloid is the quadratic and doubly ruled surface given by the Cartesian equation

$$z = \frac{y^2}{b^2} - \frac{x^2}{a^2} \tag{1}$$

(left figure). An alternative form is

$$z = x\,y \tag{2}$$

(right figure; Fischer 1986), which has parametric equations

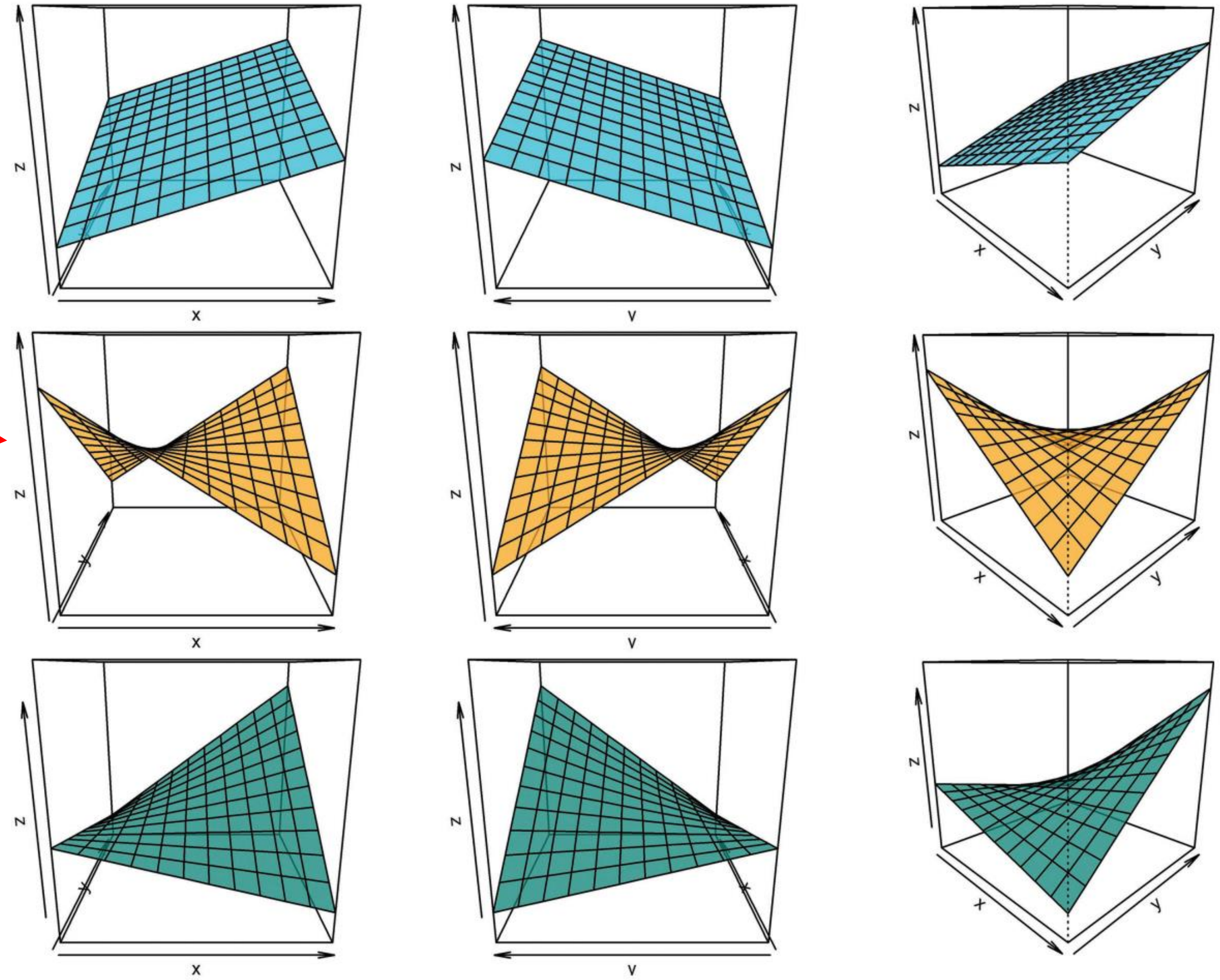
$$x(u, v) = u \tag{3}$$

$$y(u, v) = v \tag{4}$$

$$z(u, v) = u\,v \tag{5}$$

Interactions Between Quantitative Predictors

$$z = \boxed{a + (b \cdot x) + (c \cdot y)} + (d \cdot x \cdot y)$$

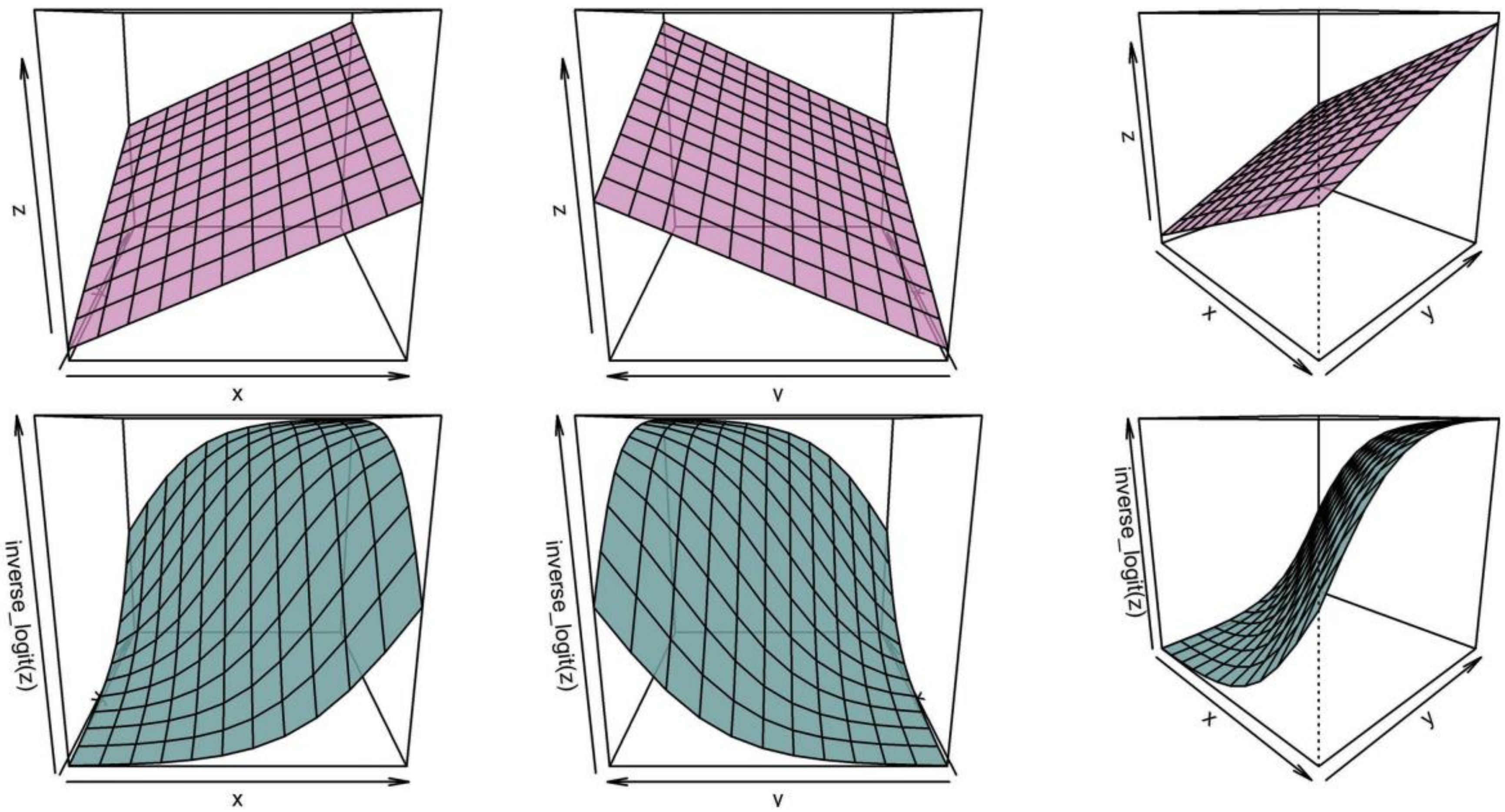


Logistic Regression with Multiple Quantitative Predictors

$$y \sim a + b \cdot x + c \cdot y$$

logits

probabilities

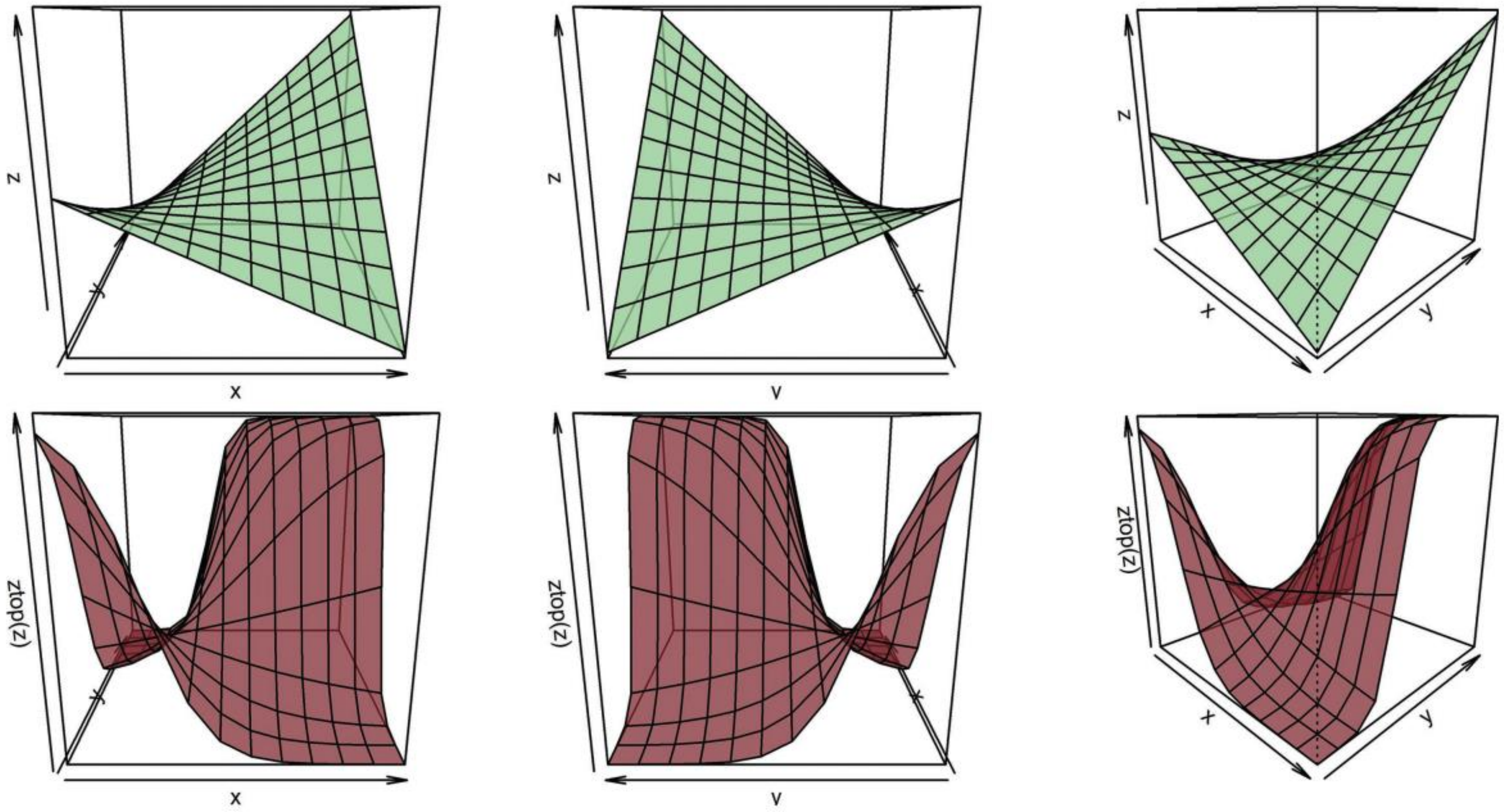


Logistic Regression with Hyperbolic Parabaloids

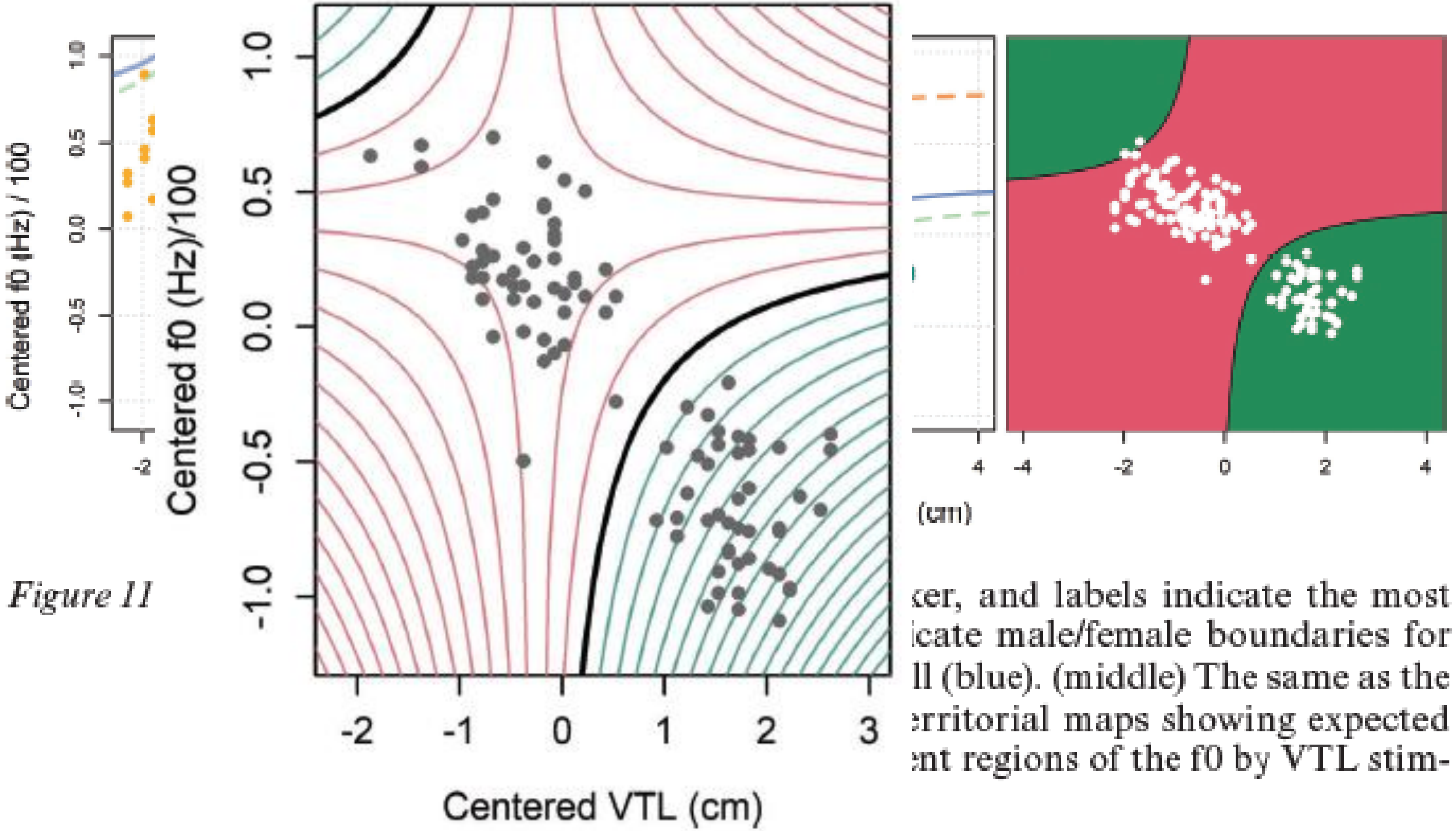
$$y \sim a + b \cdot x + c \cdot y + d \cdot x \cdot y$$

logits

probabilities



“Weird” Classification spaces



Results (Height)

- People like to see plots of the data
- Pick ones that help understand some result

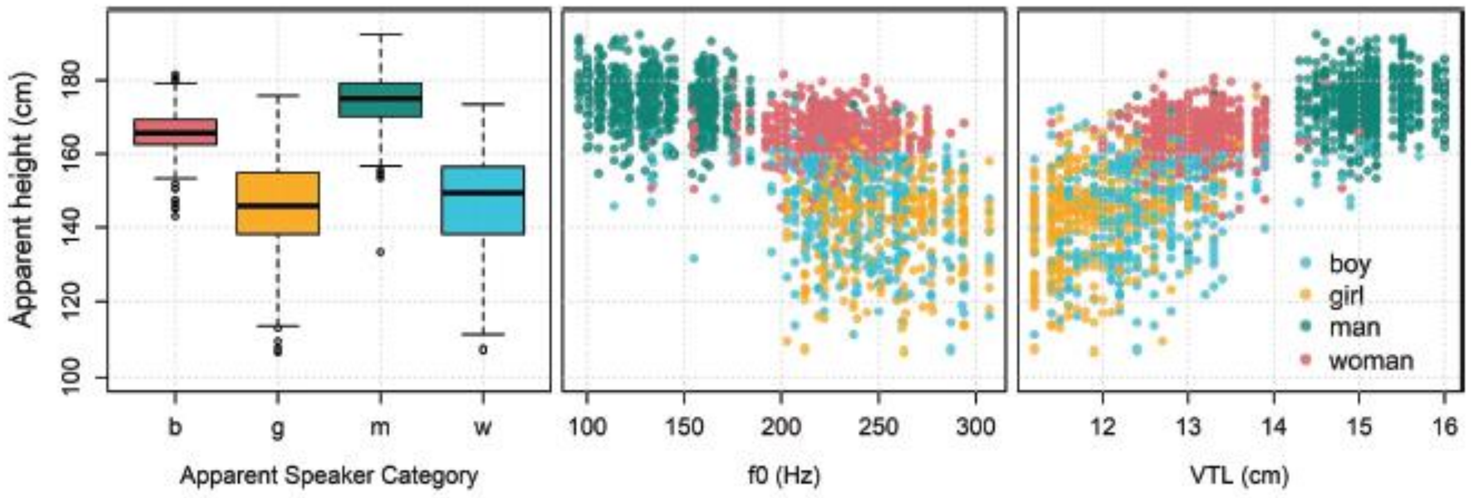


Figure 13.4 (left) Distribution of apparent height judgments across all listeners, grouped by apparent speaker category. (middle) Distribution of individual apparent height judgments based on the fundamental frequency (f0) of the stimulus. (right) Distribution of individual apparent height judgments based on the vocal-tract length (VTL). Point colors represent the modal category judgment made by listeners, for each token.

Results (Height)

- Provide summaries of all fixed effect parameters
- Use tables!!

Table 13.1 Posterior means, standard deviations, and 2.5% and 97.5% quantiles for our regression model 'fixed' effect parameter estimates

	Estimate	*Est.error*	*Q2.5*	*Q97.5*
Intercept	160.17	1.14	157.90	162.41
vtl	3.04	0.54	1.97	4.10
f0	−2.92	0.88	−4.67	−1.17
A1	7.20	1.12	5.01	9.39
G1	−0.09	0.70	−1.43	1.34
vtl:A1	−1.70	0.49	−2.64	−0.67
f0:A1	−0.46	0.78	−1.97	1.10
vtl:G1	0.47	0.42	−0.35	1.30
f0:G1	0.67	0.81	−1.00	2.20
A1:G1	−1.23	0.57	−2.32	−0.11
vtl:A1:G1	−0.70	0.37	−1.43	0.05
f0:A1:G1	0.90	0.90	−0.83	2.75

Results (Height)

- Don’t bombard readers with numbers and facts
- People need a story to remember things
- You can remind them of expected results, and research questions
- Explain your results to the reader!!

Results (Height)

- You don't need to talk about <u>everything</u>. Focus on what is likely to matter

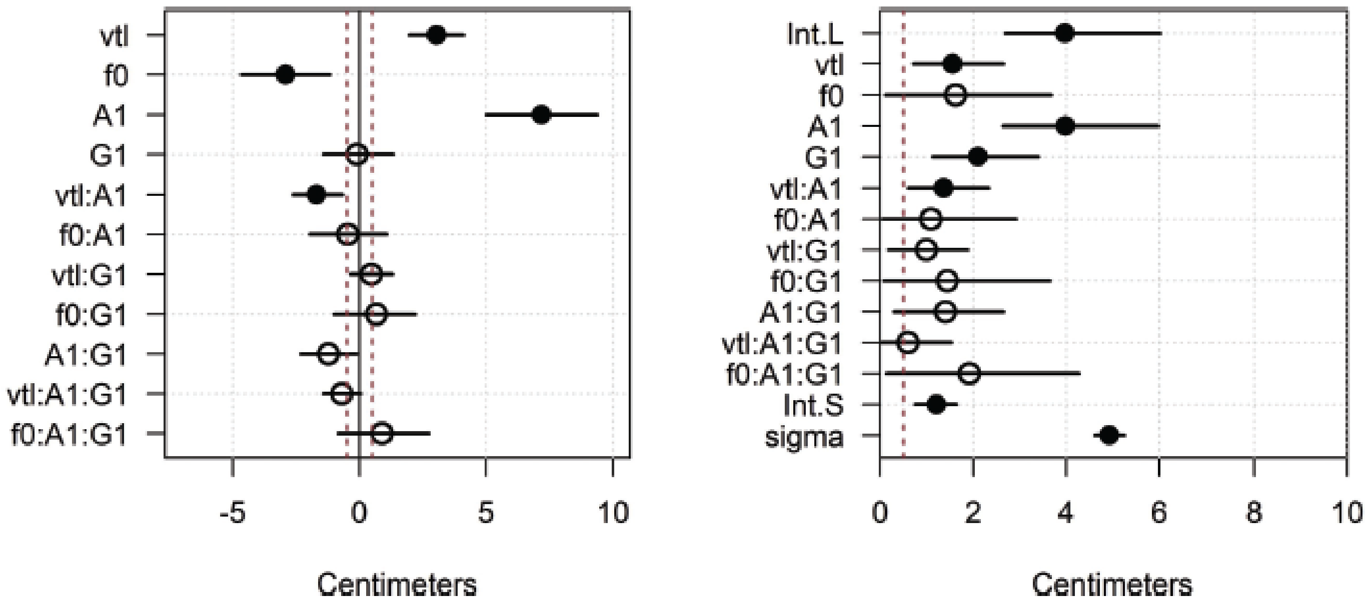


Figure 13.5 (left) Posterior distribution of model fixed effects. (right) Posterior distribution of random effect standard deviation estimates. Points indicate means, lines indicate 95% credible intervals. Dashed vertical lines indicate 0.5 cm away from 0. Points whose 95% credible intervals are at least 0.5 cm away from zero are filled.

Results (Height)

- What do these results <u>mean</u>?

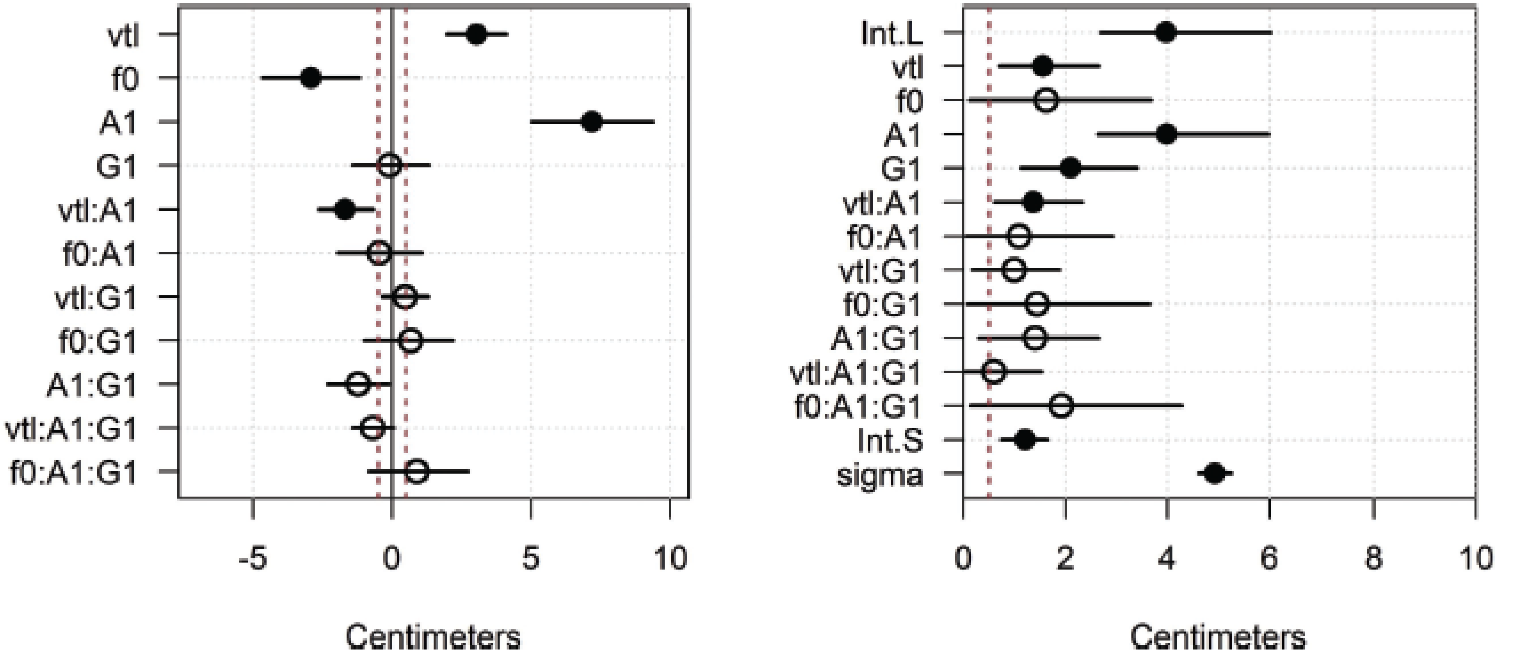


Figure 13.5 (left) Posterior distribution of model fixed effects. (right) Posterior distribution of random effect standard deviation estimates. Points indicate means, lines indicate 95% credible intervals. Dashed vertical lines indicate 0.5 cm away from 0. Points whose 95% credible intervals are at least 0.5 cm away from zero are filled.

Results (Height)

- Figure compares model parameters to data
- Unpack interaction by presenting simple effects

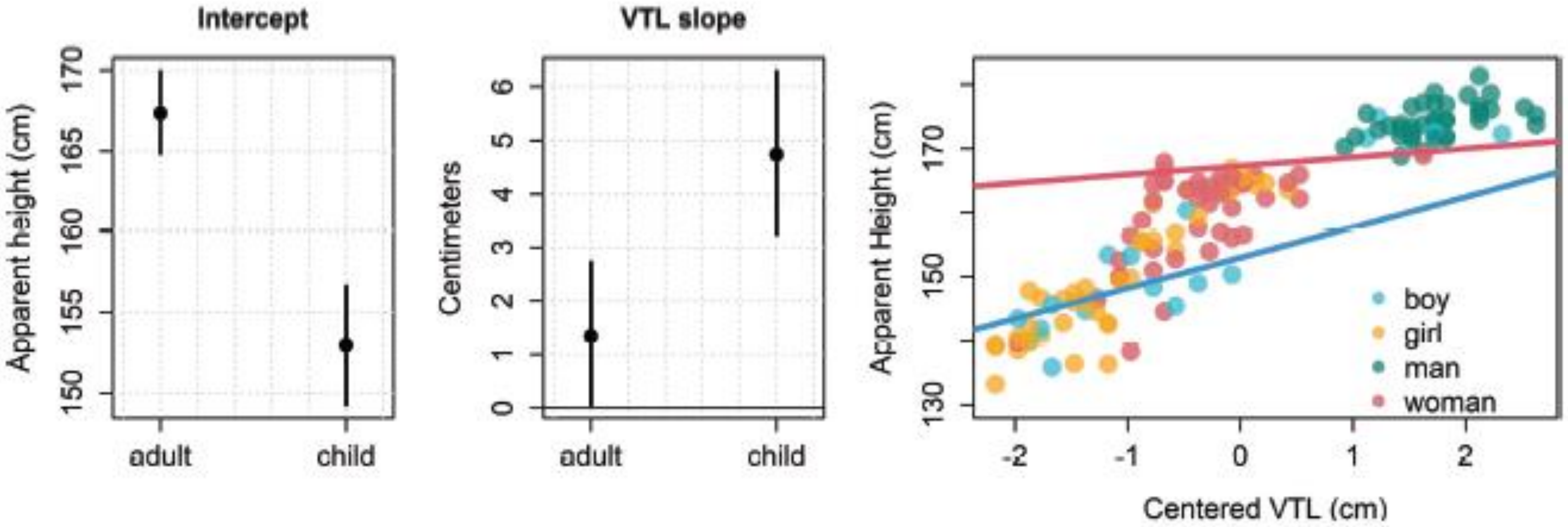


Figure 13.6 Posterior means and 95% credible intervals for model intercepts (left), and slopes (middle), of the lines relating speaker vocal-tract length (VTL) and apparent height, for apparent children and adults. (right) Average apparent height for each stimulus plotted against stimulus VTL. Point colors represent modal classifications for each speaker.

Results (Height)

- Show that model result is ‘really’ in the data if you can

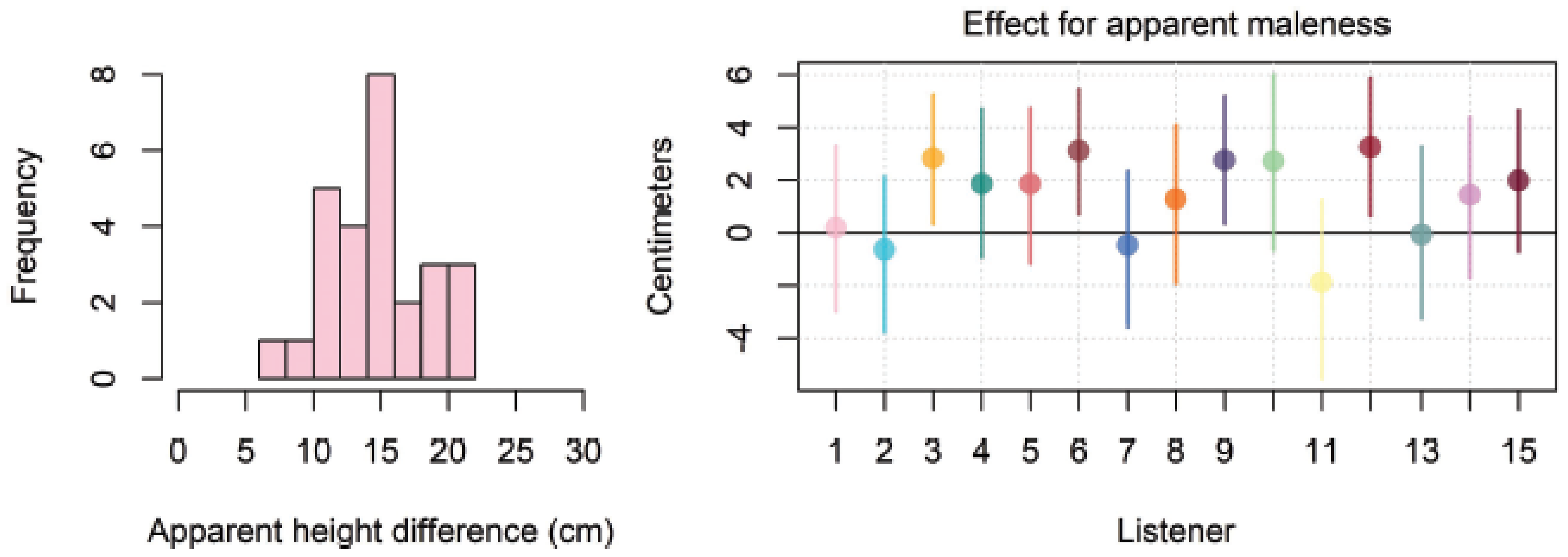
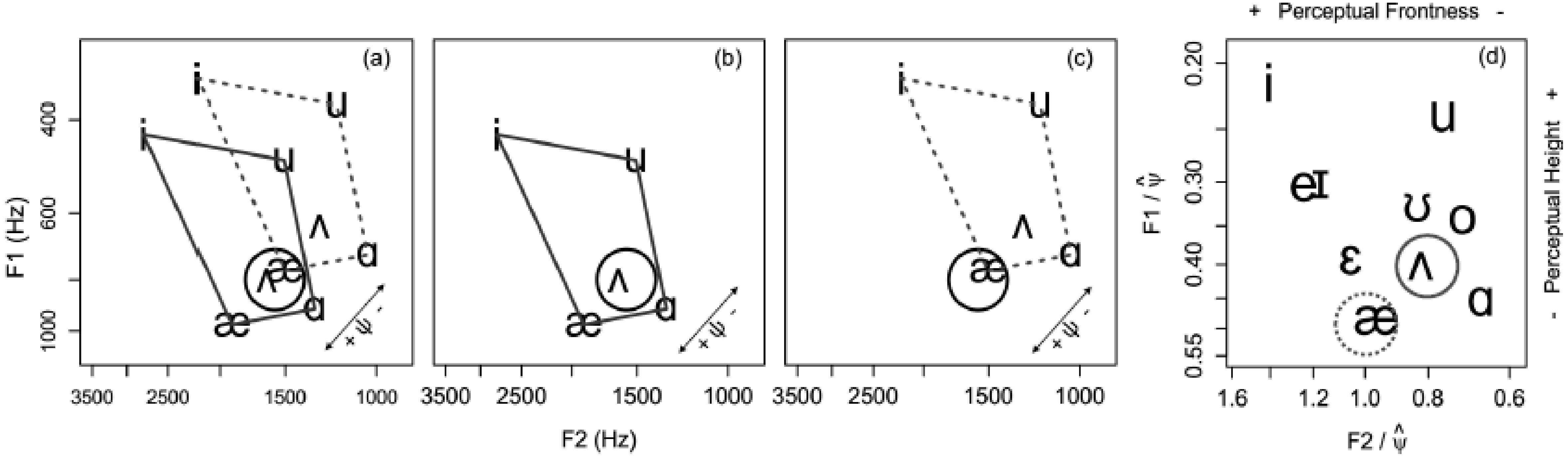


Figure 13.7 (left) Distribution of apparent height difference for 27 speakers identified as both adults and children at least five times each. The difference reflects apparent adult height minus apparent child height. (right) Posterior distribution and 95% credible intervals for listener-dependent effects for apparent maleness.

Barreda (2020)



Barreda (2020)

EFFECT	MEAN	*SD*	95% HDI		*p*-VALUE
(intercept)	−1.00	0.09	−1.16	−0.83	< 0.001
LF	4.24	0.16	3.94	4.55	< 0.001
HF	−1.47	0.09	−1.65	−1.29	< 0.001
f0	−0.80	0.11	−1.00	−0.58	< 0.001
Height	0.64	0.11	0.41	0.86	< 0.001

TABLE 3. Means, standard deviations (*SD*), 95% highest-density intervals (HDI), and *p*-values for the posterior probability of regression parameters for the vowel model. Lower formants (LF) indicate the effects of F1 and F2, while higher formants (HF) indicate the effects of formants F3 and higher.

Barreda (2020)

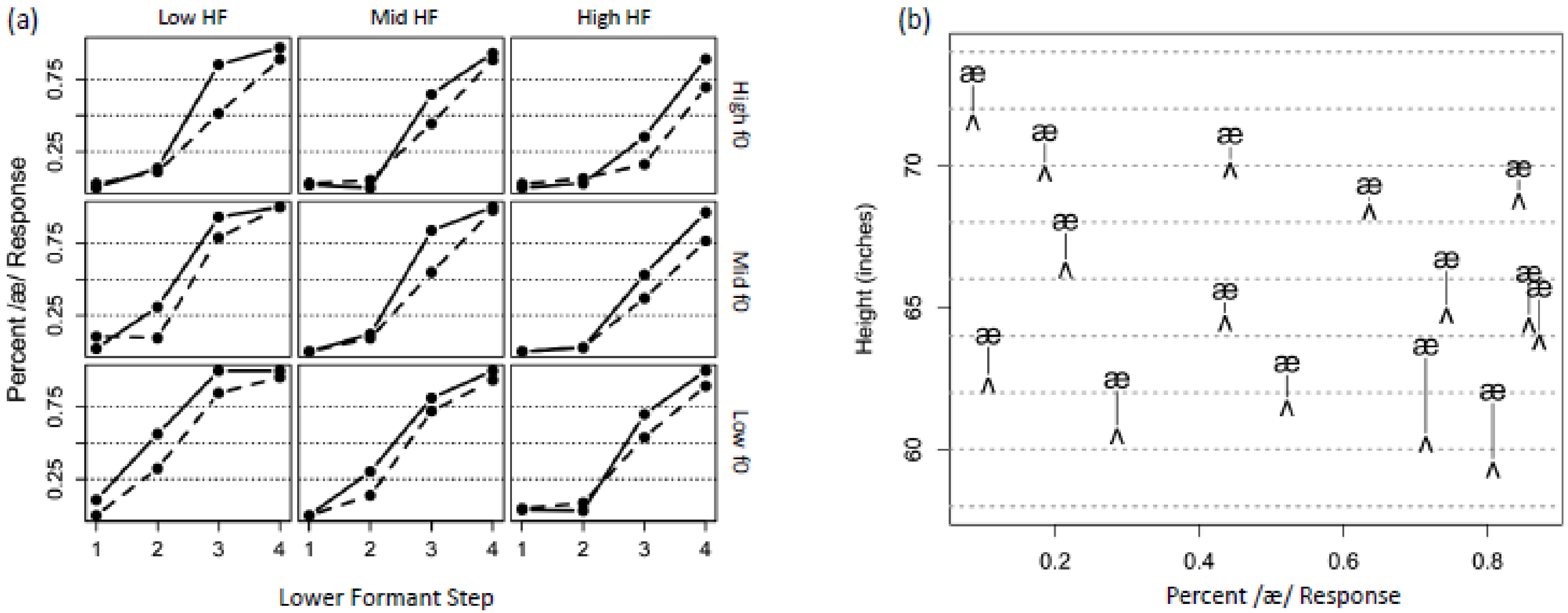


FIGURE 9. (a) Proportion of trials reported as /æ/ across all subjects for each individual stimulus. Lines indicate average responses for upper quartile of size responses (solid line) and lower quartile of size responses (dashed line). (b) Points indicate average height reported for individual tokens when these were classified as either /æ/ or /ʌ/; lines indicate the difference in apparent height for each token according to phonemic classification. Stimuli are arranged on the x-axis based on the rate at which they were classified as /æ/. Only stimuli that were identified as both categories at least ten times each are presented.

Results (Gender)

- Still plot dichotomous data!
- Show proportions (or logits?) if its binary

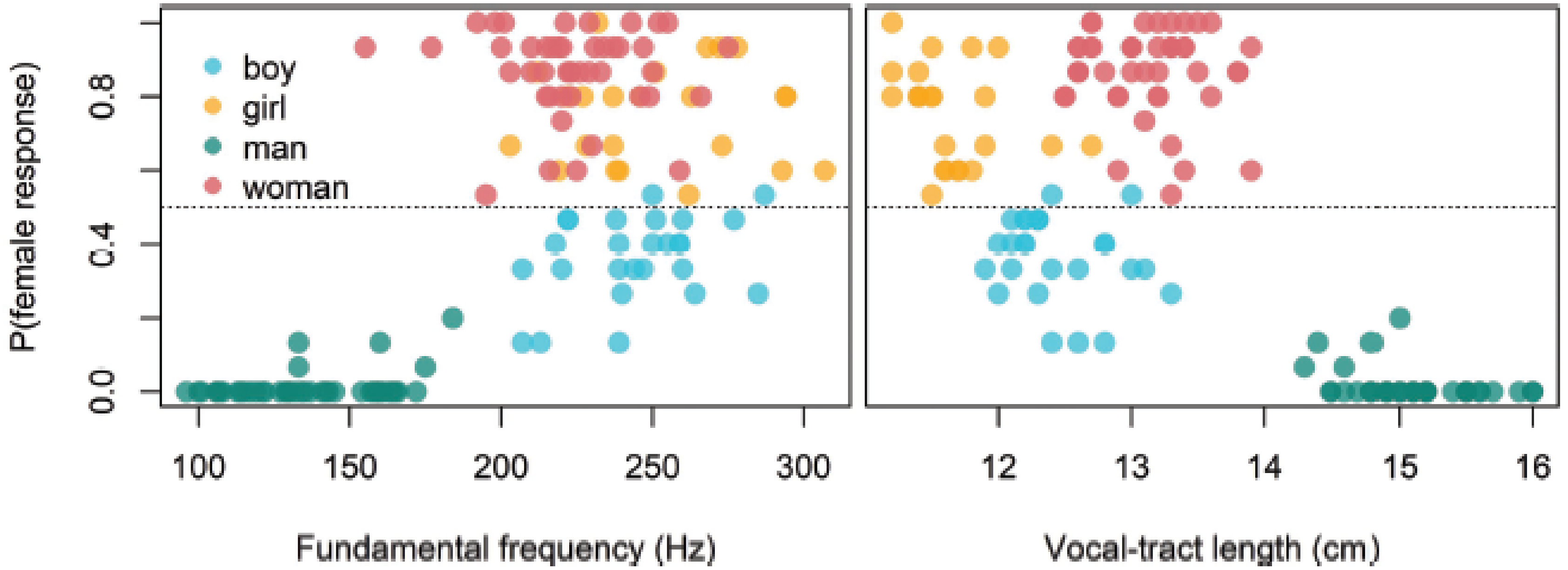


Figure 13.8 Probability of observing a female response for each speaker, across all listeners. Point color reflects modal speaker classification. Points are organized according to speaker fundamental frequency (left) and speaker (acoustic) vocal-tract length (right).

Results (Gender)

- What's going on here?

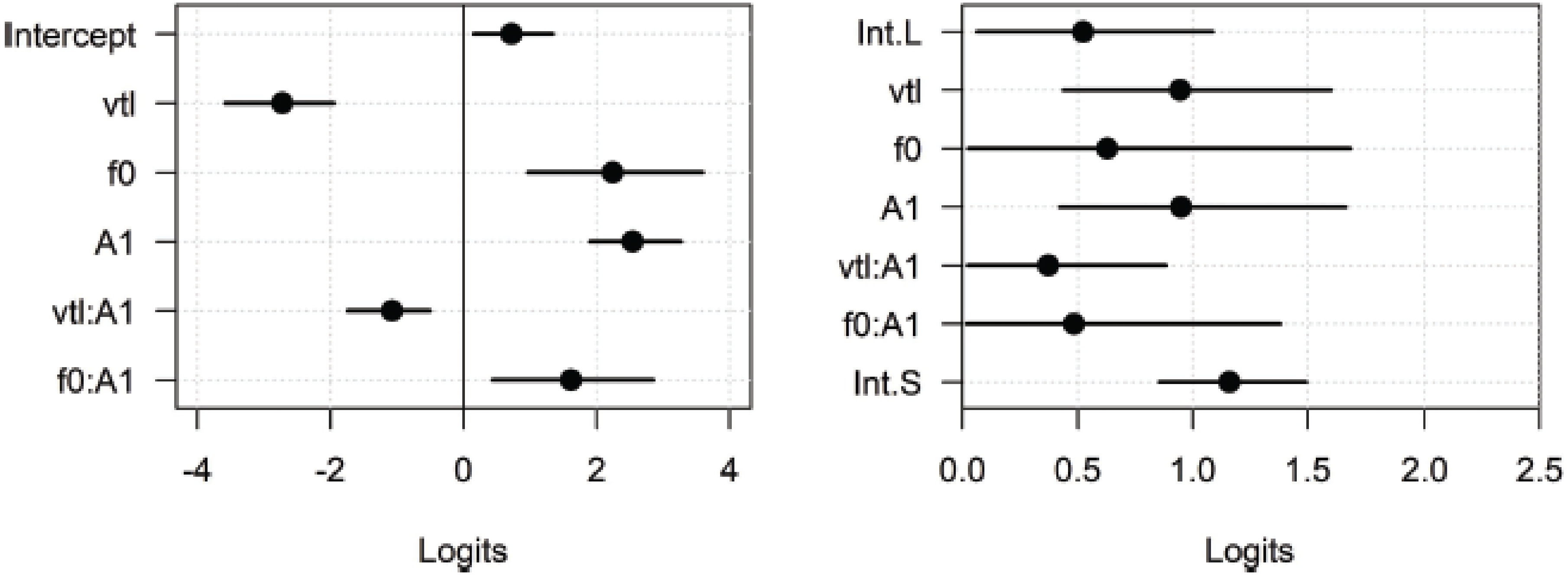


Figure 13.9 (left) Posterior distribution of model fixed effects. (right) Posterior distribution of random effect superpopulation standard deviation estimates. Points indicate means, lines indicate 95% credible intervals.

Discussion

- Talk about how readers should interpret the results

13.7.1 Effect of apparent age on the perception of femaleness

Apparent adultness increased the probability of observing a female response and also increased the effects of VTL and f0, leading to steeper slopes along each dimension. For example, the interaction between f0 and apparent age (`f0:A1`) has the effect of reducing the effect of f0 to nearly zero for children (`f0 + (-A1:f0)`, mean = 0.62, s.d. = 0.71, 95% C.I = [−0.77, 2.08]), and nearly doubling it for adults (`f0 + A1:f0`, mean = 3.85, s.d. = 1.07, 95% C.I = [1.86, 5.97]). The joint effects of f0 and VTL, and the way that these vary in relation to apparent age can be considered using territorial maps, as in Figure 13.10. Territorial maps present category boundaries in a stimulus space and indicate which categorical outcome is most probable in each region of the space.

Discussion

- Some figures help you talk about results more than directly present them
- These figures can go in results or discussion

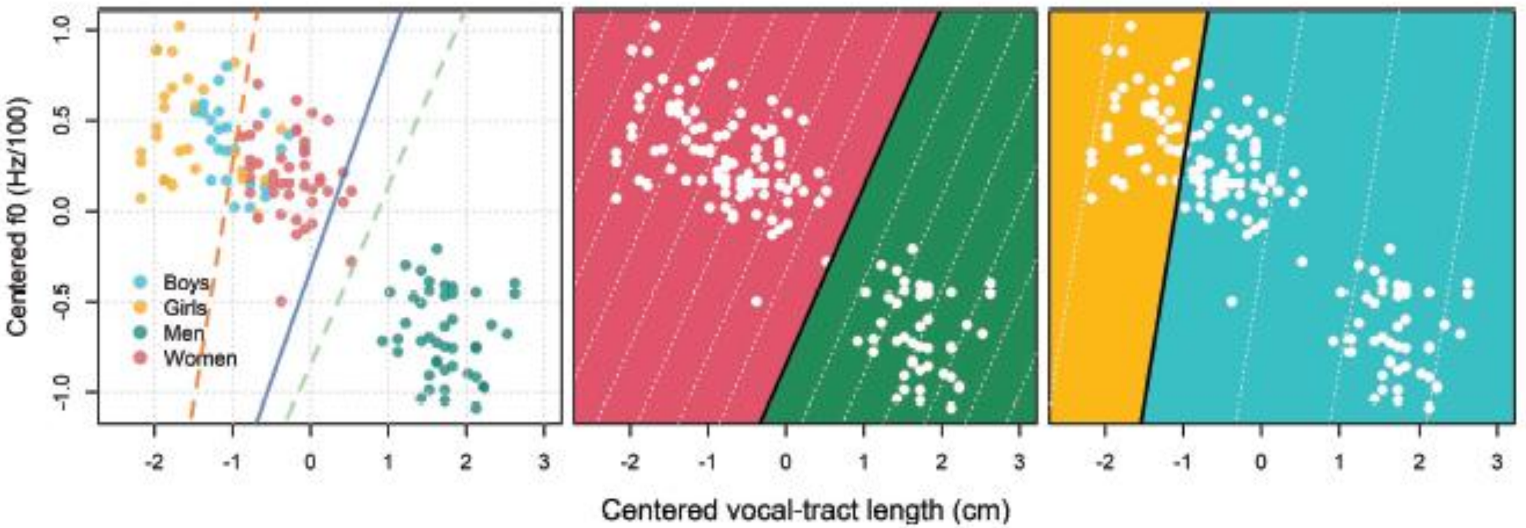


Figure 13.10 (left) Each point represents a single speaker, and labels indicate the most common group classification. Lines indicate male/female boundaries for adults (green), children (orange), and overall (blue) implied by our model. (middle) Territorial map showing expected classifications for apparent adults in different regions of the f0 by VTL stimulus space. (right) Same as the middle, but for apparent children. In each territorial map, dotted lines indicate an increase/decrease of 2 logits in the expected value of observing a female response, starting at a value of zero for the solid black line.

Discussion: Individual Differences™

- This plot compares individual listener data and behavior

- Listener/speaker variation is important in many fields

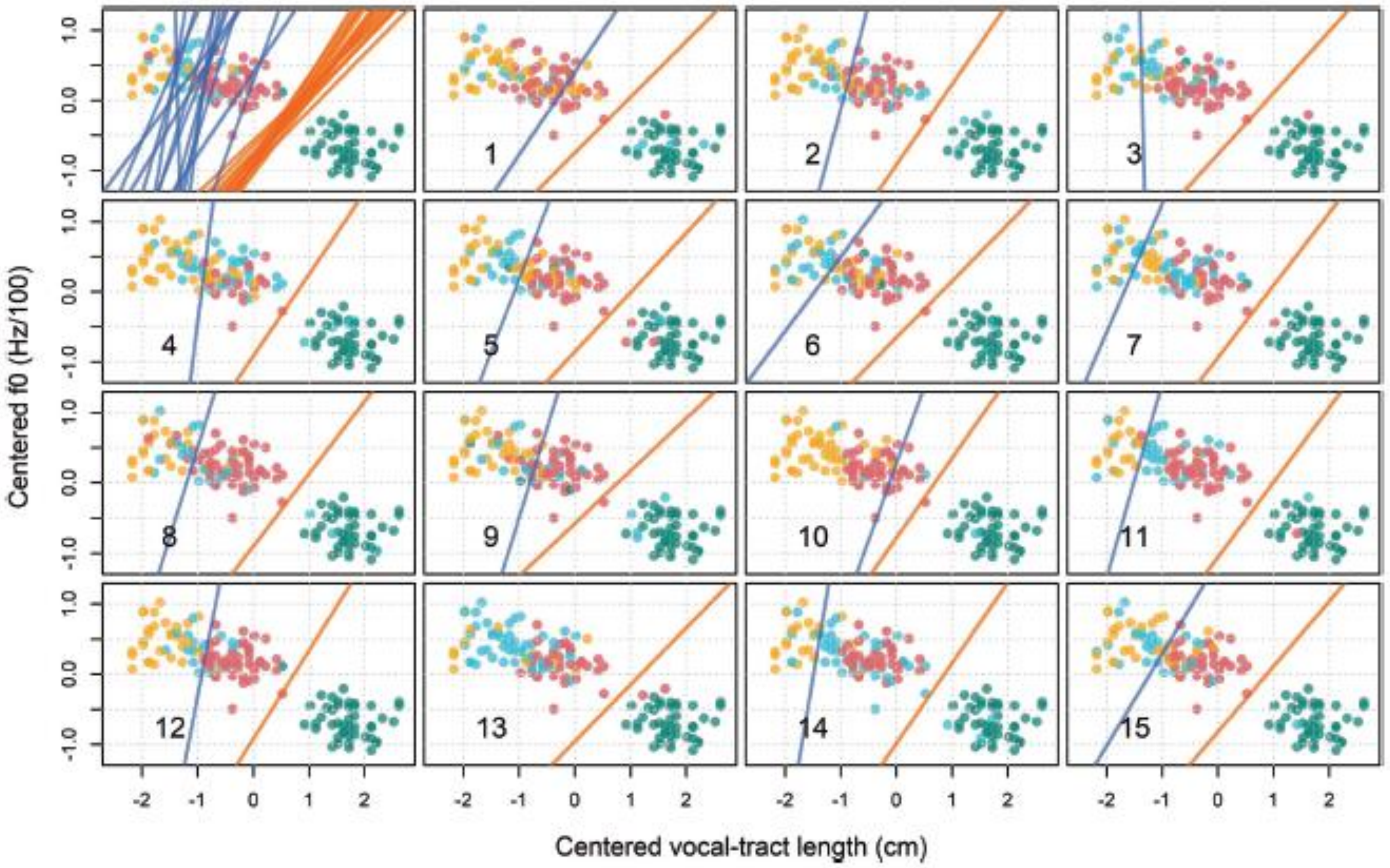


Figure 13.11 The first plot compares all listener-dependent category boundaries between men/women (orange) and boys/girls (blue). Subsequent plots present each listener's individual boundaries. Point colors indicate the listener's individual speaker classifications. The number on each plot indicates the listener number.

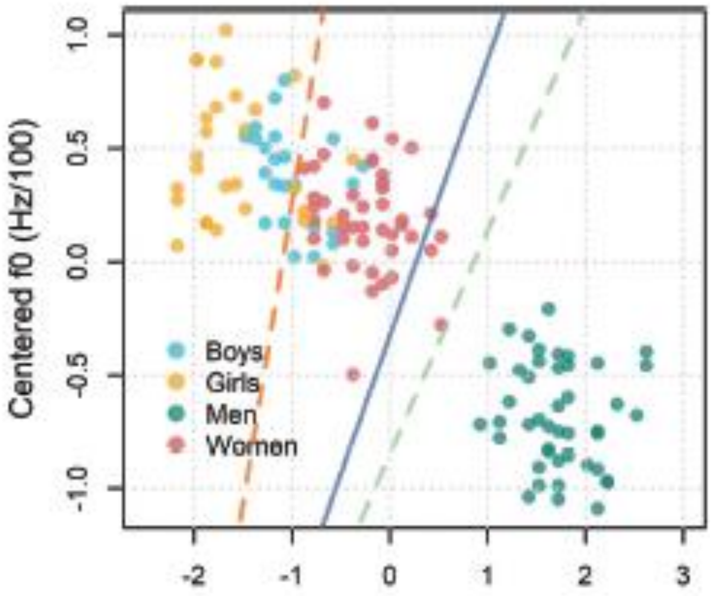
Discussion: Remaining Questions

- “As we can see in Figure 13.3, boys and women overlap substantially in their gross acoustics. As a result, to the extent that their gender is identified correctly, we might expect that this would be on the basis of ‘something else’. [...] What could the ‘something else’ aiding in the classification of boys and adult women be? One candidate is what is known as *prosodic* information, that is, information about timing, rhythm, pitch movement, and so on. It’s also possible that other more subtle acoustic information is involved, for example, information about the breathiness or creakiness of the voice. In any case, results highlight the fact that the communication and identification of gender from speech is more complicated than what can be explained by speaker f0 and VTL.”

Discussion: Remaining Questions

- We conclude boys are identified as boys based on “something else”

- Discussing the unresolved mysteries makes work more interesting



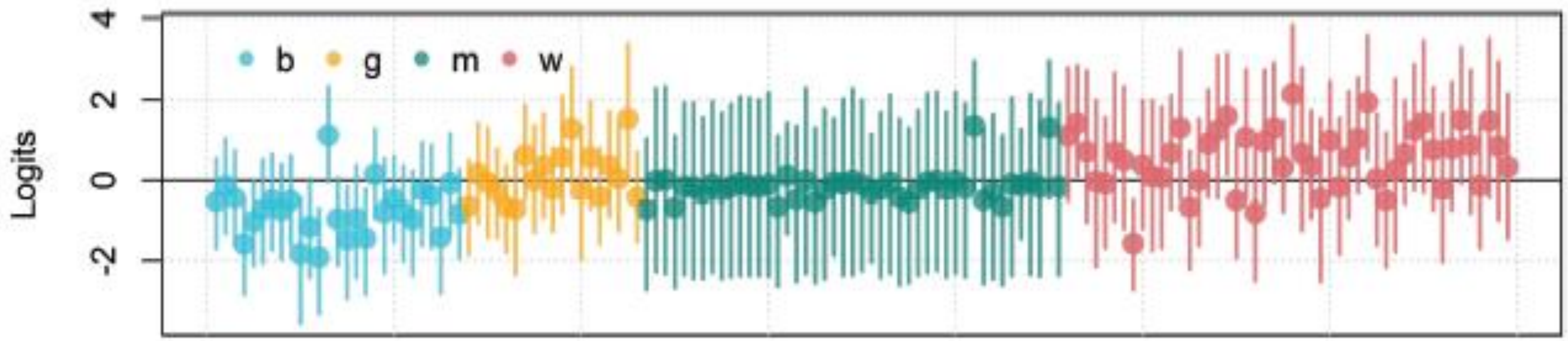


Figure 13.12 Speaker random effects. Point colors indicate the veridical speaker category for boys (b), girls (g), men (m), and women (w).

Should You fit the Model?

- As we progress, you *can* fit more and more complicated models.
- But *should* you?
- Two things to worry about:
  - Identifiability: the ability to estimate unique, independent values for all your parameters.
  - Support: Having enough data to realistically estimate all your parameters.

Collinearity

- A set of predictor vectors is linearly independent when there is no vector of non-zero numbers that can be used to combine our predictors such that they <u>always</u> equal zero.

$$0 = x_1 \cdot a_1 + x_2 \cdot a_2 + \ldots + x_n \cdot a_n$$

If this is possible the x predictors are not linearly dependent.

Collinearity

- You cannot fit a model using height in meters and height in centimeters and estimate both effects independently.

```
exp_data$vtl_m = exp_data$vtl / 100
```

```
exp_data$vtl + exp_data$vtl_m * -100
```

- These predictors are linearly dependent!

$$x_1 \cdot a_1 + x_2 \cdot a_2 = 0$$
$$\mathrm{vtl} \cdot 1 + \mathrm{vtl_m} \cdot -100 = 0$$

Collinearity

```r
model_bad_1 =
  brms::brm (height ~ vtl_m + vtl, data = exp_data, chains = 4, cores = 4,
       warmup = 1000, iter = 3500, thin = 2,
       prior = c(brms::set_prior("normal(176, 50)", class = "Intercept"),
                 brms::set_prior("normal(0, 15)", class = "sigma")))
```

```
## Population-Level Effects:
##           Estimate Est.Error   l-95% CI  u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept    45.69      2.37      41.17     50.47 1.00     4031     4425
## vtl_m      1383.91 246800.93 -531293.80 343951.54 2.12        5       13
## vtl          -5.28   2468.01   -3430.88   5321.56 2.12        5       13
```

Semi-Collinearity

- Linear dependence is binary.

- Correlation is a continuous measure of linear dependence (basically).

```
cor (exp_data$vtl_m, exp_data$vtl)
## [1] 1
```

- What if we make our predictors almost linearly dependent?

```
set.seed(1)
exp_data$vtl_m_noise = exp_data$vtl_m +
  rnorm (length(exp_data$vtl_m),0,sd(exp_data$vtl_m)/10)

cor (exp_data$vtl, exp_data$vtl_m_noise)
## [1] 0.9946
```

Semi-Collinearity: Not that Bad

```r
model_bad_2 =
  brms::brm (height ~ vtl_m_noise + vtl, data = exp_data, chains = 4,
           cores = 4,warmup = 1000, iter = 3500, thin = 2, prior = priors)
```

```
## Population-Level Effects:
##             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      45.75      2.36    41.16    50.35 1.00     4523     4087
## vtl_m_noise  -201.86    175.11  -544.11   139.03 1.00     3504     3697
## vtl            10.57      1.76     7.14    14.01 1.00     3439     3866
```

Semi-Collinearity: Could be Better

```r
model_good =
  brms::brm (height ~ vtl, data = exp_data, chains = 4, cores = 4,
       warmup = 1000, iter = 3500, thin = 2,
       prior = c(brms::set_prior("normal(176, 50)", class = "Intercept"),
                 brms::set_prior("normal(0, 15)", class = "sigma")))
```

```
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept    45.72      2.38    41.04    50.40 1.00     5052     4679
## vtl           8.55      0.18     8.21     8.90 1.00     5072     4831
```

```r
model_bad_2 =
  brms::brm (height ~ vtl_m_noise + vtl, data = exp_data, chains = 4,
         cores = 4,warmup = 1000, iter = 3500, thin = 2, prior = priors)
```

```
## Population-Level Effects:
##             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      45.75      2.36    41.16    50.35 1.00     4523     4087
## vtl_m_noise  -201.86    175.11  -544.11   139.03 1.00     3504     3697
## vtl            10.57      1.76     7.14    14.01 1.00     3439     3866
```

Predictable Values for Predictors

- You can't fit models where the value of one categorical predictors can be guessed based on the values of other predictors.

- This is again a problem of linear dependence.

- This is why we can't estimate both levels of a two-group factor.

```
x = cbind (intercept=rep(1,4), x1=rep(c(1,0),2),
x2=rep(c(0,1),2))
x
##      intercept x1 x2
## [1,]         1  1  0
## [2,]         1  0  1
## [3,]         1  1  0
## [4,]         1  0  1
```

```
x[,1] + x[,2]*(-1) + x[,3]*(-1)
## [1] 0 0 0 0
```

Predictable Values for Predictors

- And why we can't estimate all 4 levels of a 4-level factor.

```
x = cbind (intercept=rep(1,4), C1=c(1,0,0,0), C2=c(0,1,0,0),
           C3=c(0,0,1,0),C4=c(0,0,0,1))
X
##      intercept C1 C2 C3 C4
## [1,]         1  1  0  0  0
## [2,]         1  0  1  0  0
## [3,]         1  0  0  1  0
## [4,]         1  0  0  0  1
```

```
x[,1] + x[,2]*(-1) + x[,3]*(-1) + x[,4]*(-1) + x[,5]*(-1)
## [1] 0 0 0 0
```

Predictable Values for Predictors

- And why we can't include group, age, *and* gender.

```
x = cbind (intercept=rep(1,4), C1=c(1,0,0,0), C2=c(0,1,0,0),
           C3=c(0,0,1,0),A1=c(0,0,1,1), G1=c(0,1,0,1),A1G1=c(1,0,0,1))
```

```
x
##      intercept C1 C2 C3 A1 G1 A1G1
## [1,]         1  1  0  0  0  0    1
## [2,]         1  0  1  0  0  1    0
## [3,]         1  0  0  1  1  0    0
## [4,]         1  0  0  0  1  1    1
```

```
x[,1]*1 + x[,2]*(-1) + x[,3]*(-1) + x[,5]*(-1)
## [1] 0 0 0 0
```

```
x[,1]*1 + x[,3]*(-1) + x[,4]*(-1) + x[,7]*(-1)
## [1] 0 0 0 0
```

Predictable Values for Predictors

```
model_bad_3 =
  brms::brm (height ~ C + A*G, data = exp_data, chains = 4, cores = 4,
       warmup = 1000, iter = 3500, thin = 2,
       prior = c(brms::set_prior("normal(176, 15)", class = "Intercept"),
                 brms::set_prior("normal(0, 15)", class = "sigma")))
```

```
## Population-Level Effects:
##            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept    157.98      0.20   157.57   158.35 1.06       61       78
## C1          1066.31   2604.72 -4431.79  5158.68 2.07        5       18
## C2           496.91   1751.32 -2543.19  3708.83 1.87        6       13
## C3            74.39   2086.76 -4113.31  2801.74 1.99        5       21
## A1           793.68   1363.24 -2383.78  3019.96 1.90        6       12
## G1           567.63   2171.69 -4107.90  3364.23 2.04        5       18
## A1:G1        283.89   1112.60 -2295.30  2141.73 2.26        5       18
```

This is not ideal.

Saturated Models

- Saturated models have one parameter for every observation.
- Without shrinkage this means that there is no random variation in the model, i.e., the error cannot be estimated.

```
exp_data$S = factor (exp_data$S)
exp_data$L = factor (exp_data$L)

model_bad_4  =
  brms::brm (height ~ S*L, data = exp_data, chains = 4, cores = 4,
       warmup = 1000, iter = 3500, thin = 2,
       prior = c(brms::set_prior("normal(176, 15)", class = "Intercept"),
                 brms::set_prior("normal(0, 15)", class = "sigma")))
```

This model crashes my session of R.

Nearly-Saturated Models

- A model can be 'nearly'-saturated.
- What if you have 2 observations per speaker per listener. You can fit this model (probably), but should you?
- Follow up: What is your n?
- Think of your n for each individual parameter rather than overall.