EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Analyzing time series data using GAMs

Jacolien van Rij & Martijn Wieling | LSA 2015, Chicago

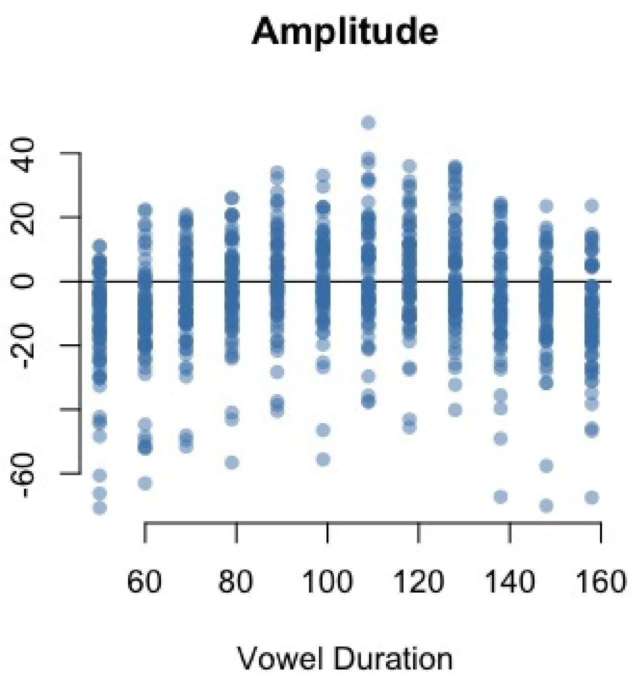
Day 1 • July 7

1) Short summary of **linear regression**
2) Intro **GAMM**s
3) Example of GAMM analysis: **articulography data**

1. Linear regression

Example data

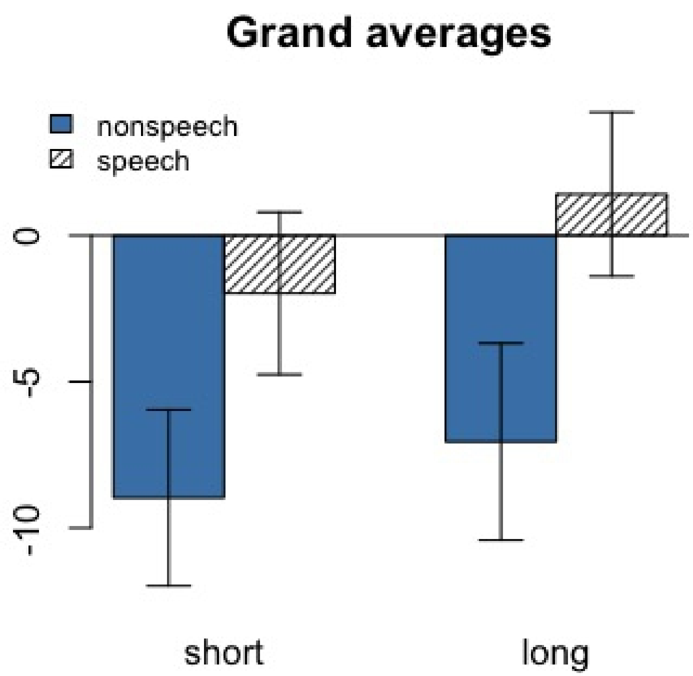
- MEG data of Tomaschek et al. (2013)
  - *Vowel duration* and *mode*

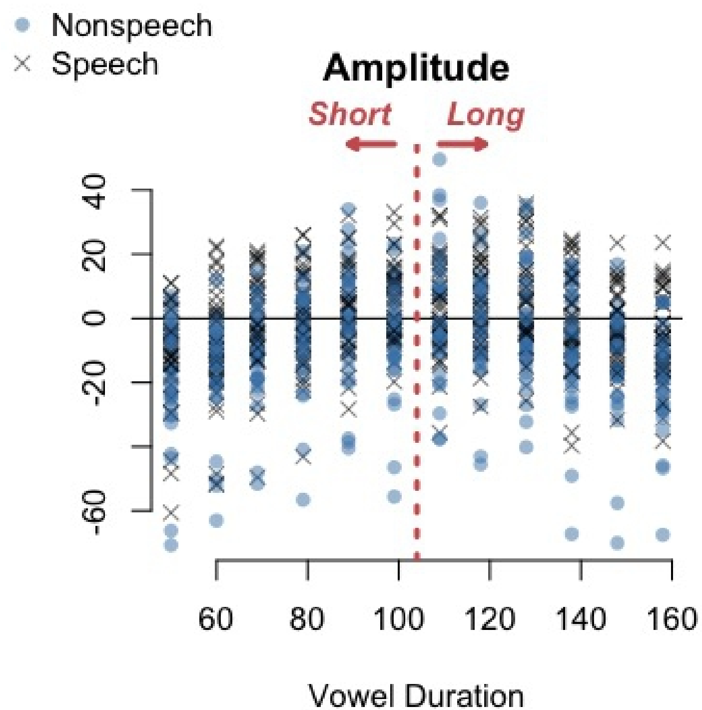


NIMH Image library, http://infocenter.nimh.nih.gov/il/public_il/image_details.cfm?id=80

ANOVA

- Categorical data



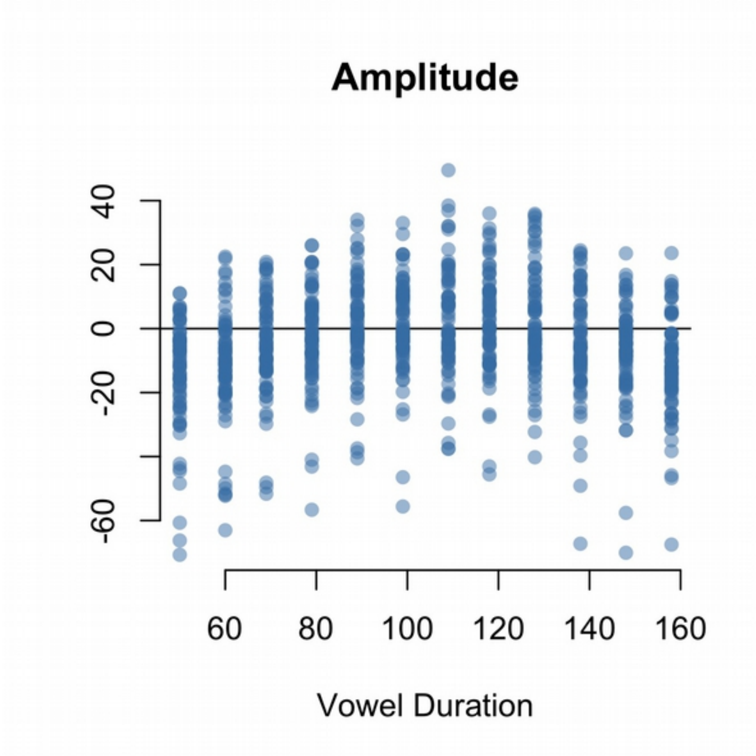


Linear regression

- Continuous and categorical predictors

- Example 1:
Could the value of *Vowel Duration* explain the measured *Amplitude*?

    - Vowel duration = numeric

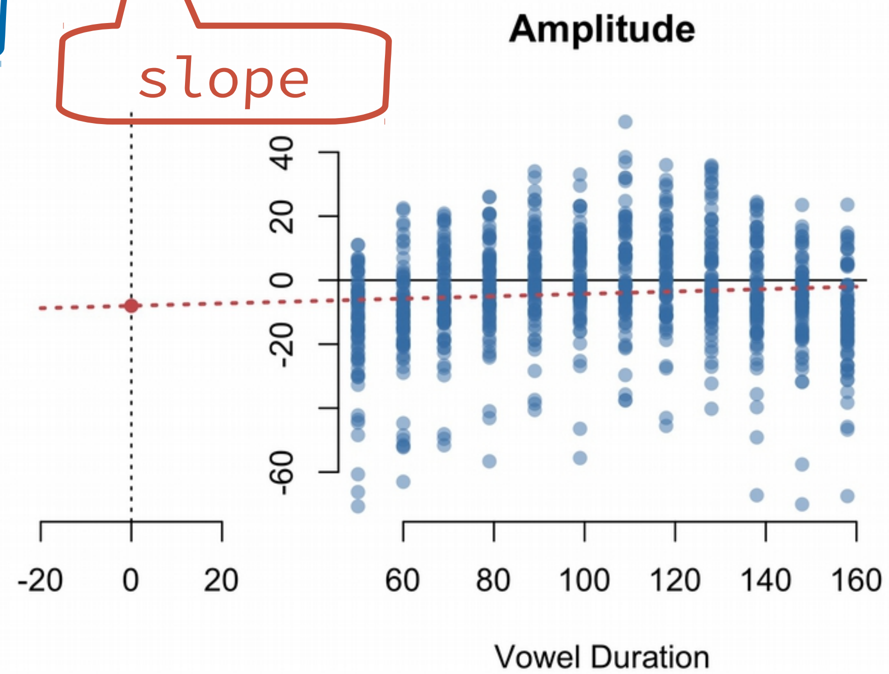


Continuous predictors

- Regression formula: $y = \beta_0 + \beta_{VD} * x + \epsilon$

intercept

slope



Continuous predictors

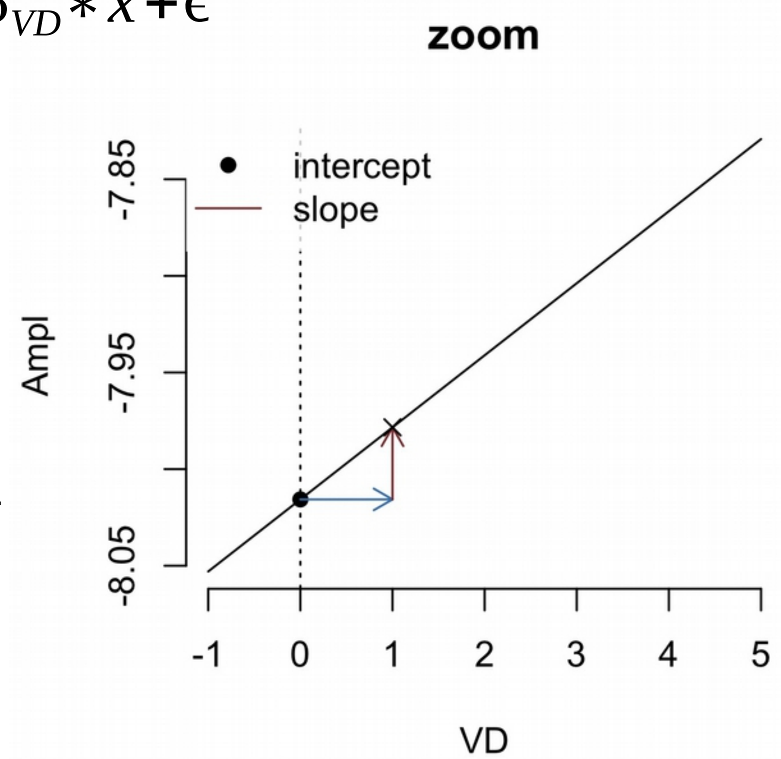
- Regression formula: $y = \beta_0 + \beta_{VD} * x + \epsilon$

```
lm1 <- lm(Amp ~ VD, data=dat)
summary(lm1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.01569    1.78739  -4.485 8.24e-06
VD           0.03730    0.01638   2.277    0.023
```

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

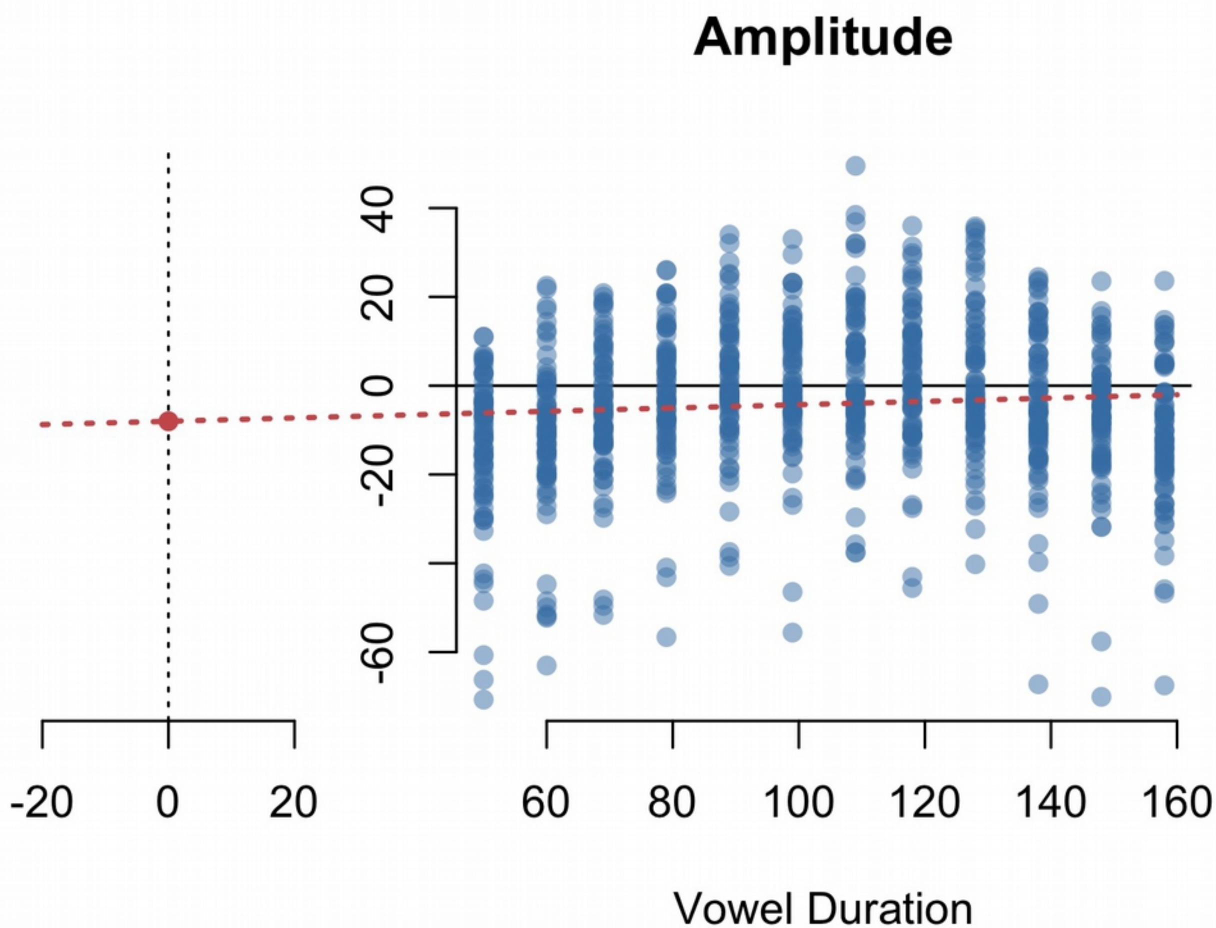
Intercept and slope

- Regression formula: $y = \beta_0 + \beta_{VD} * x + \epsilon$



- $ß_0$ =Intercept: value of y when x=0 (crossing with y-axis)
- $ß_1$ =Slope: change in y-value with with increasing 1 unit on x-axis
- Y = -8.02+0.03730*VD

Intercept and slope



Categorical predictors

- *Mode*: “speech”, “nonspeech”

- Regression formula: $y = \beta_0 + \beta_{VD} * x_{VD} + \beta_{Mode} * x_{Mode} + \epsilon$

- Contrast coding

```
contrasts(dat$Mode)
     sp
ns   0
sp   1
```

Categorical predictors

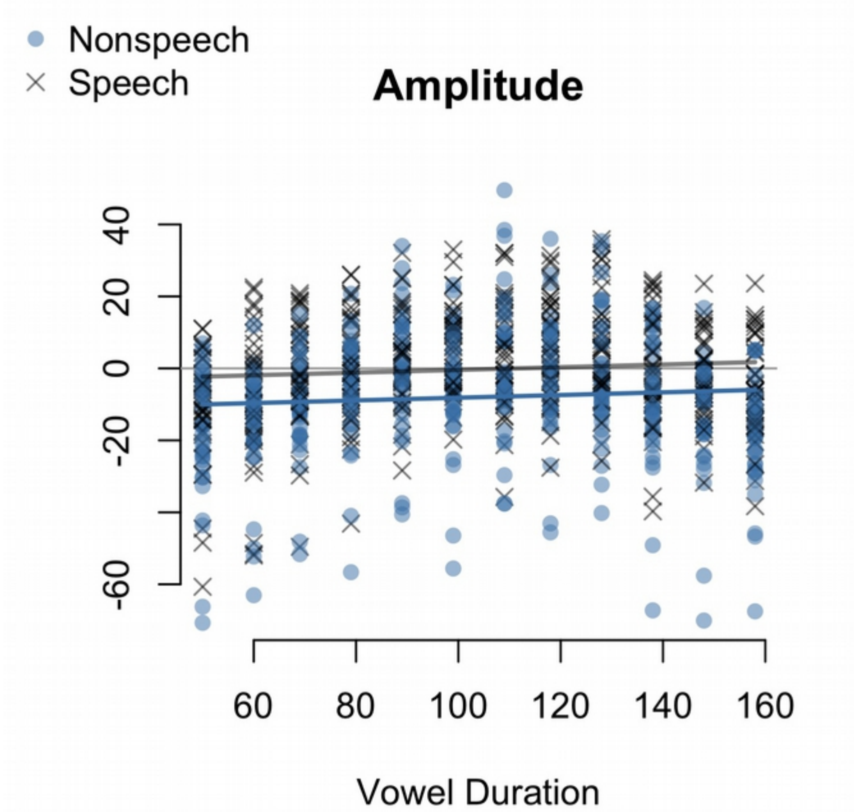
- Regression formula: $y = \beta_0 + \beta_{VD} * x_{VD} + \beta_{Mode} * x_{Mode} + \epsilon$

```
lm2 <- lm(Amp ~ VD+Mode, data=dat)
summary(lm2)

Coefficients:
              Estimate Std. Error t

(Intercept) -11.87865    1.82184
VD            0.03730    0.01594
Modesp        7.72592    1.07965
```

if Mode == “ns”:
x_{Mode} = 0
\+ 0
if Mode s==”sp”:
x_{Mode} = 1
\+ 7.72

Categorical predictor

- Two regression lines:
  - Same effect of duration,
  - but difference in average amplitude between the two conditions:



Interaction

```
lm3 <- lm(Ampl ~ VD*Mode, data=dat)
#   == lm(Ampl ~ VD + Mode + VD:Mode, dat)

summary(lm3)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.87585    2.46014  -4.014 6.45e-05
VD           0.01799    0.02254   0.798    0.425
Modesp       3.72032    3.47916   1.069    0.285
VD:Modesp    0.03861    0.03188   1.211    0.226
```

Interaction

- Testing for significance:

```
anova(lm2, lm3)
Analysis of Variance Table

Model 1: Ampl ~ VD + Mode
Model 2: Ampl ~ VD * Mode
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    909 241581
2    908 241192  1     389.6 1.4667 0.2262
```

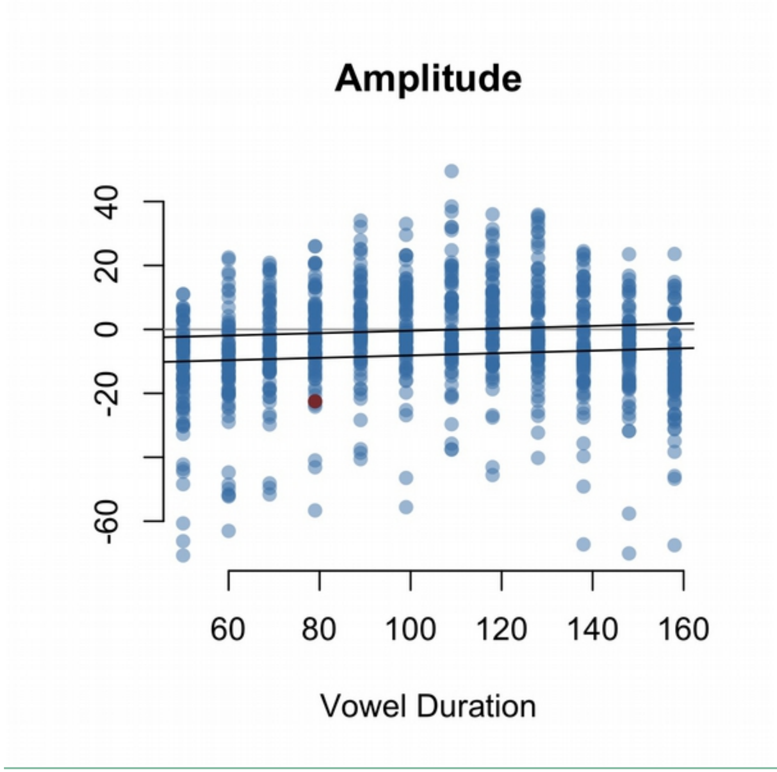
Fitted values and residuals

- Regression formula: $\boxed{y} = \beta_0 + \beta_{VD} * x_{VD} + \beta_{Mode} * x_{Mode} + \boxed{\epsilon}$

- Regression model tries to predict how the predictors influence the values of response variable
  - Fitted values: explained part of data
  - Residuals: unexplained part of data
    - Should be gaussian distributed $\epsilon \sim N(0, \sigma)$
    - No structure in residuals

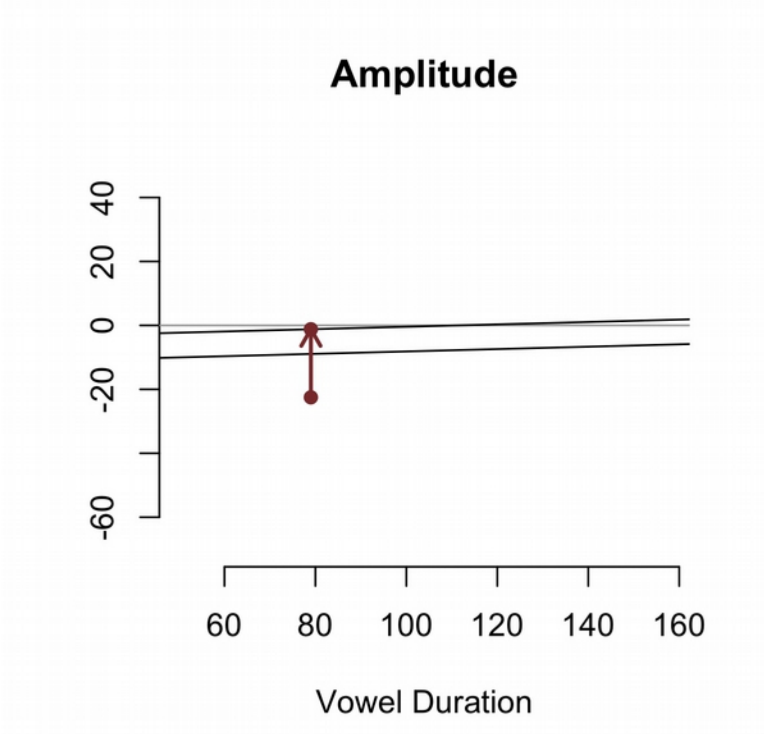
Model criticism

- Inspection of residuals: Comparing the fitted values with the actual data



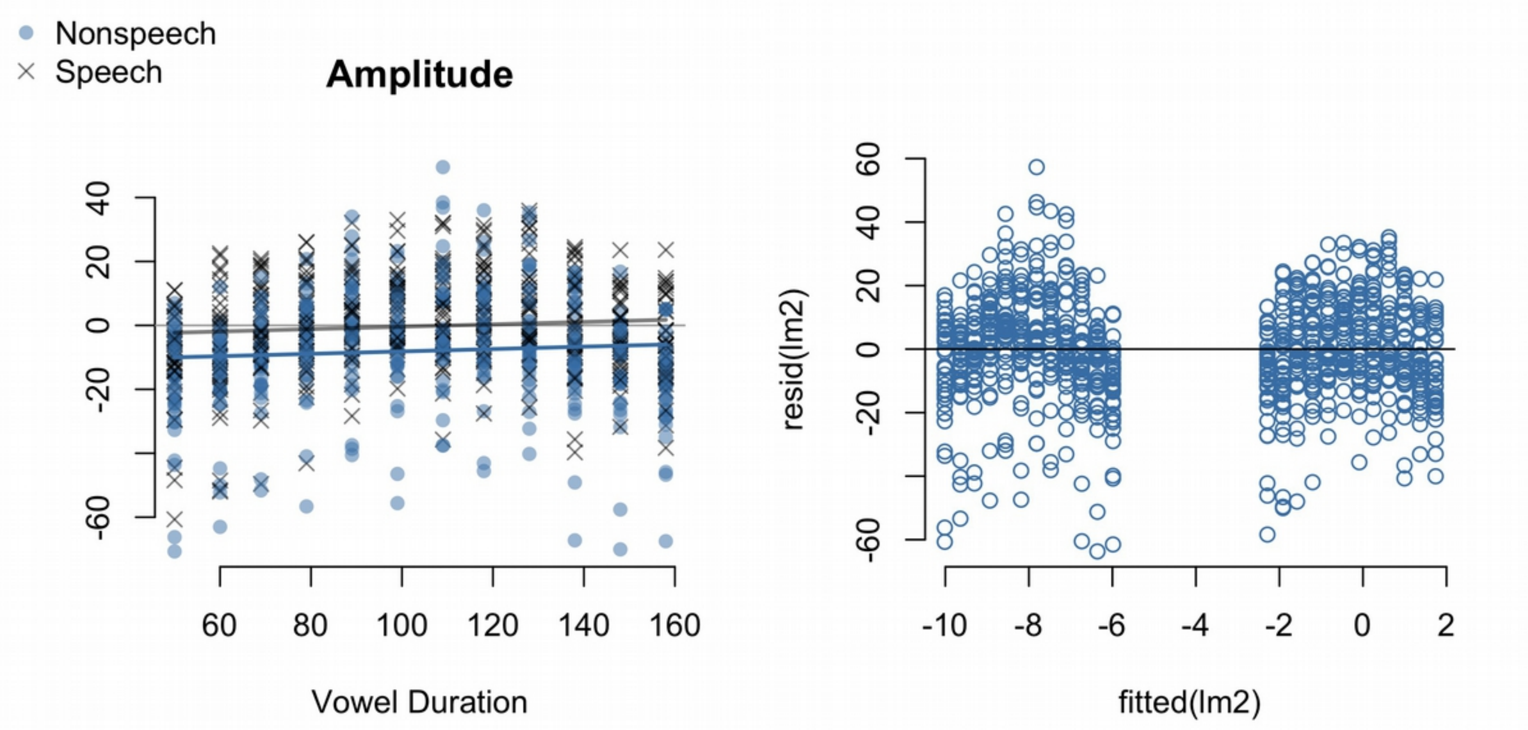
Model criticism

- Inspection of residuals: Comparing the fitted values with the actual data



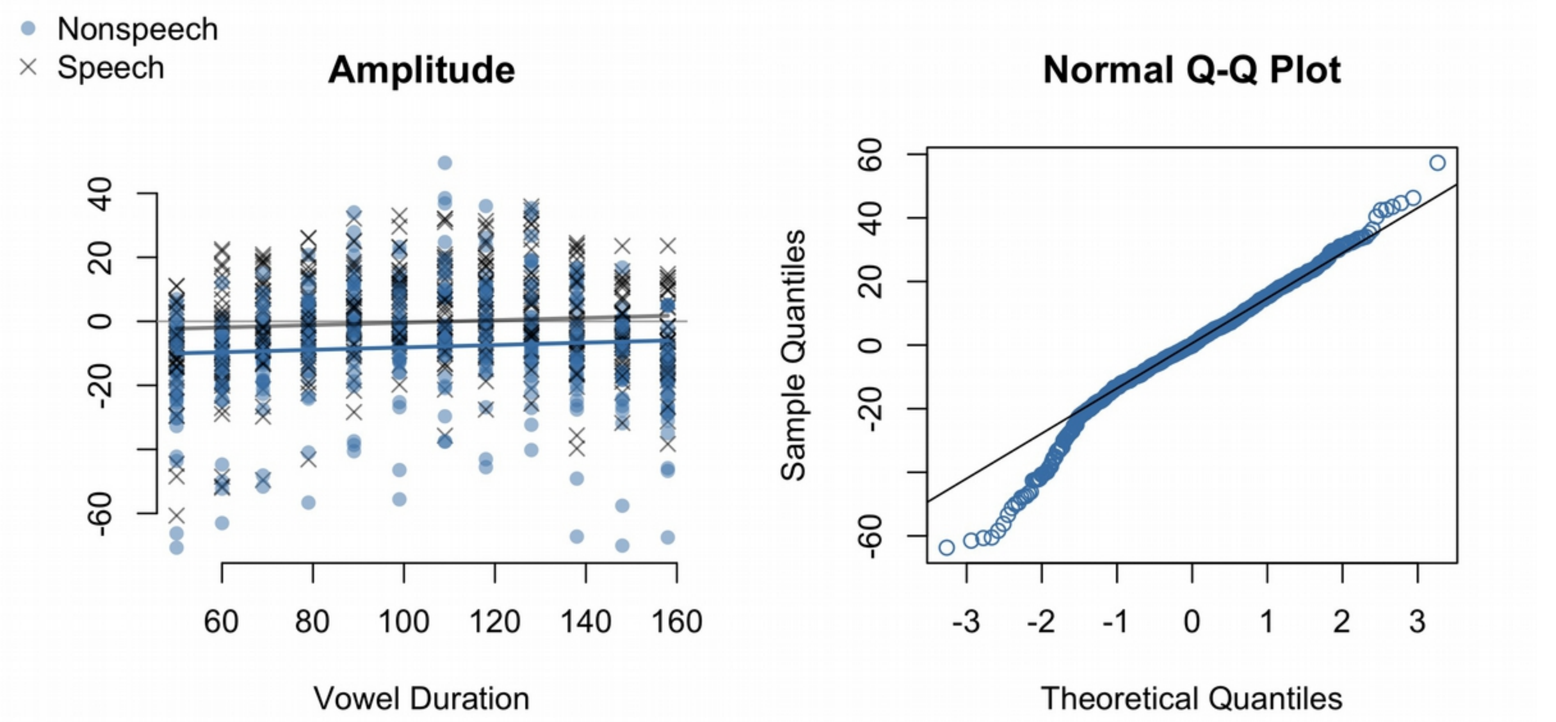
Model criticism

- Inspection of residuals: Comparing the fitted values with the actual data



Model criticism

- Inspection of residuals: Comparing the fitted values with the actual data



2. GAMMs

Generalized Additive Models (GAM)

- Relaxing assumption of linear relation between dependent variable and predictor

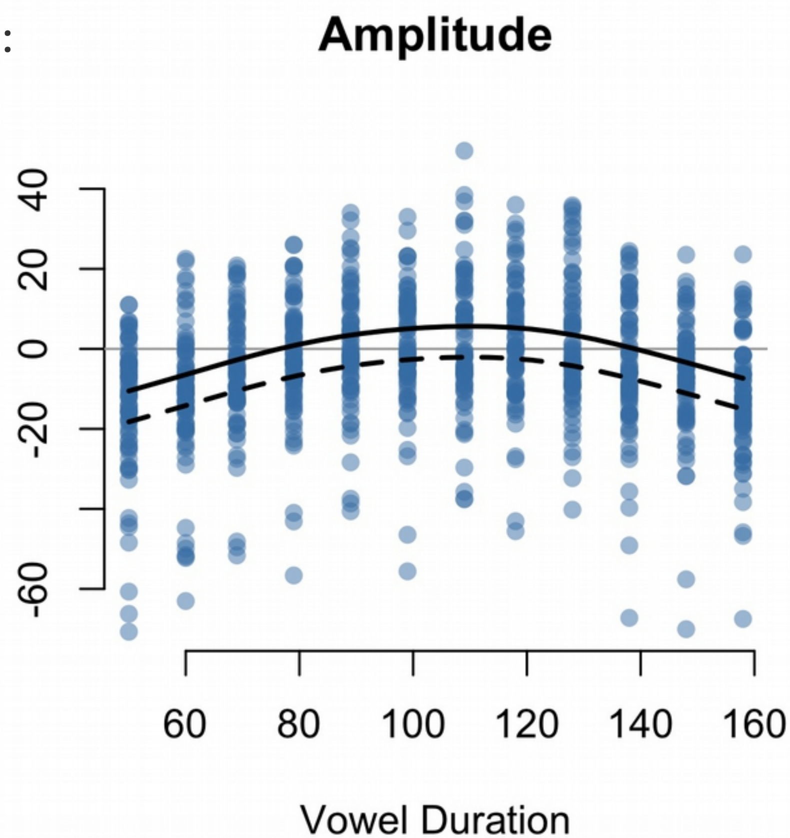
- Regression formula: $y = \beta_0 + \beta_{VD} * x_{VD} + \beta_{Mode} * x_{Mode} + \epsilon$

  becomes: $y = \beta_0 + f(x_{VD}) + f(x_{Mode}) + \epsilon$

- References:
  - GAMs (Lin & Zao, 1999)
  - R package mgcv (Wood, 2006; 2011)

Generalized Additive Models (GAM)

- Nonlinear regression lines:



3. Articulography data