Development of Brazilian Portuguese (L3) vowels by an Argentinean learner: A Bayesian approach

Ronaldo Lima Jr. (UFC/CNPq – Brazil)

Ubiratã Kickhöfel Alves (UFRGS/CNPq – Brazil)

Introduction

In the last decade, Complex Dynamic Systems Theory (CDST) has challenged researchers to try out new methods and techniques. These new approaches have allowed us to rethink the research goals in CDST studies. As a result, not only new means of data collection, but also different statistical procedures, have been employed so as to model language changes in time.

Introduction

- As we consider the development of vowel systems, rather than only asking whether new vowel categories have been fully formed, we should also enquire about the dynamic fluctuations and interactions among these vowel categories;
- With regard to these interactions, we assume that the emergence of a new vowel category may affect all of the other (pre-existing) categories in the developing system;
- Brazilian Portuguese (BP) has seven vowels in stressed position (/i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/). This poses a challenge to L1-Spanish learners, who must develop new categories for the mid-low vowels /ɛ/ and /ɔ/;
- As these new vowel categories start to emerge, it is expected that changes in all NNL vowel categories should occur, as the whole phonetic space should adapt to these new changes.

In this study...

- We investigate the development of BP vowels by an Argentinean learner living in Brazil (BP was his third language);
- This is part of a larger project aiming to investigate the interconnectedness of the learners' three languages. In this paper, we will concentrate on the L3 system only.
- The acoustic data (F1 and F2 values) from these recordings received different statistical treatments in previous studies, such as Monte-Carlo Analyses (Van Dijk, Verspoor, and Lowie 2011) and Change-Point Analyses (Taylor, 2000).
- In this study, we reanalyzed the data using a Bayesian GAM (generalized additive model), aiming to verify possible advantages of this approach to statistical inference.

Research Question

To what extent can a Bayesian multilevel GAM depict the dynamic fluctuations inherent to speech development over time?

Method

The Argentinean participant (Buenos Aires) was 36 years old and had been working as a professor in Brazil (Porto Alegre) for three years and 7 months at the time the data collections started.

He showed an A2 level of proficiency in English and a “Highly Advanced” level in Brazilian Portuguese (CELPE-BRAS exam – the official proficiency exam of BP in Brazil);

He took part in 24 data collection sessions, every 15 days (from Oct 2018 to Sept 2019);

Pronunciation instruction (following ALVES; BRISOLARA; PEROZZO, 2017’s manual) on consonant and vowel sounds was provided from data collections 10 to 15. There were 12 weekly one-to-one classes.

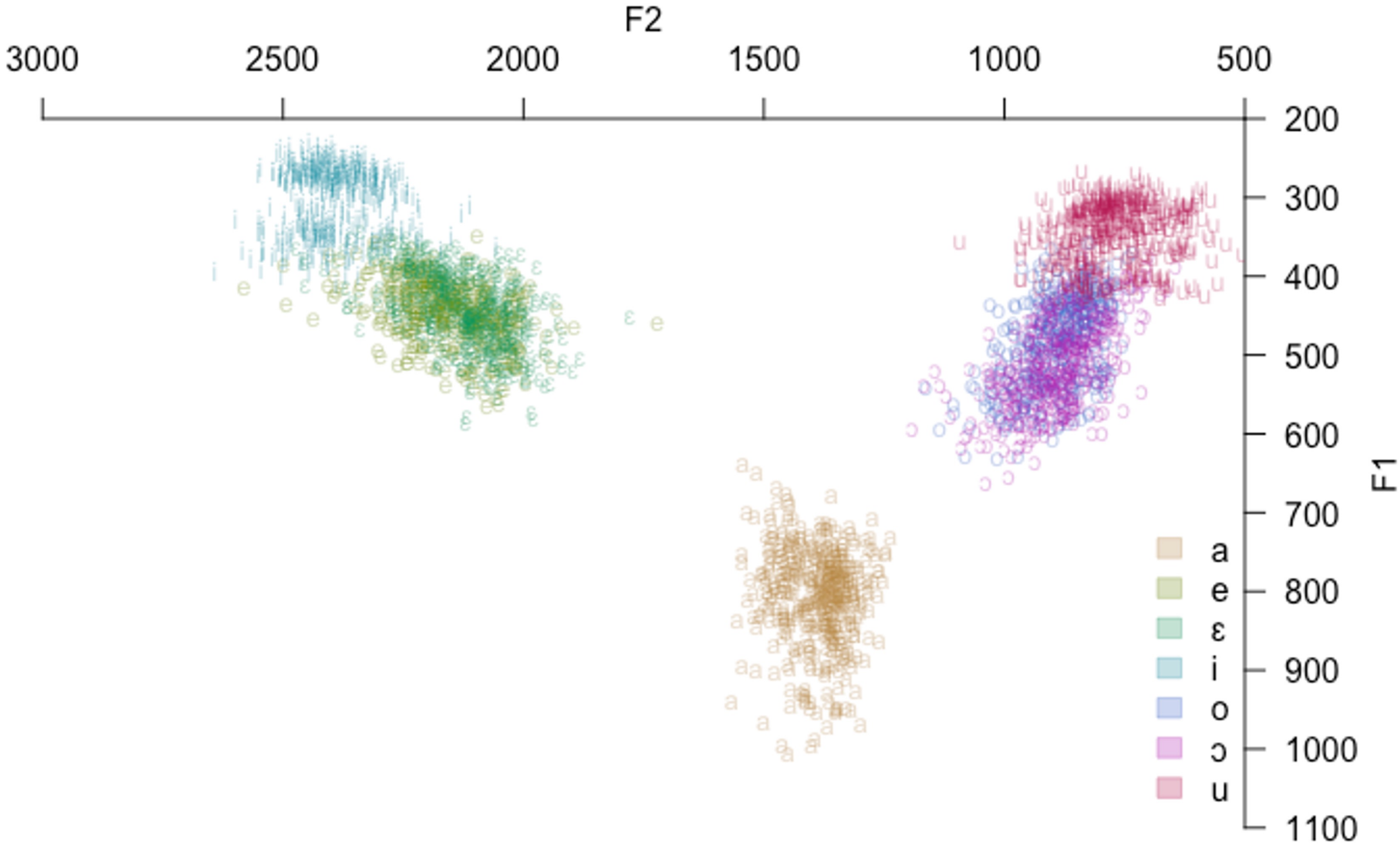
Method

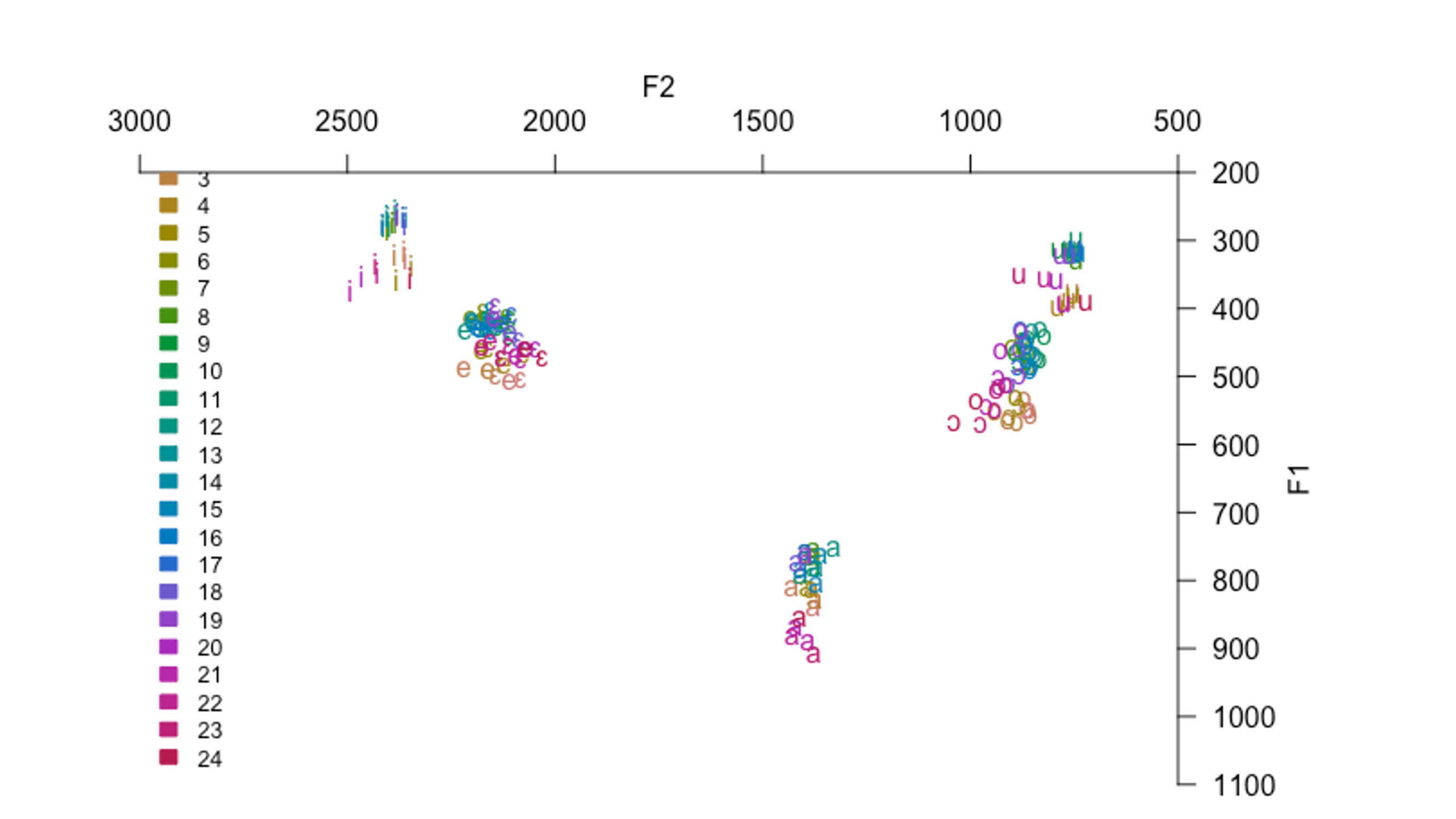
Reading instrument: the same one applied in Pereyron (2017): (carrier sentences: “Diga____”);

Acoustic analyses of F1 and F2 values (*Praat* – v. 6.1.16, Boersma & Weenink, 2019);

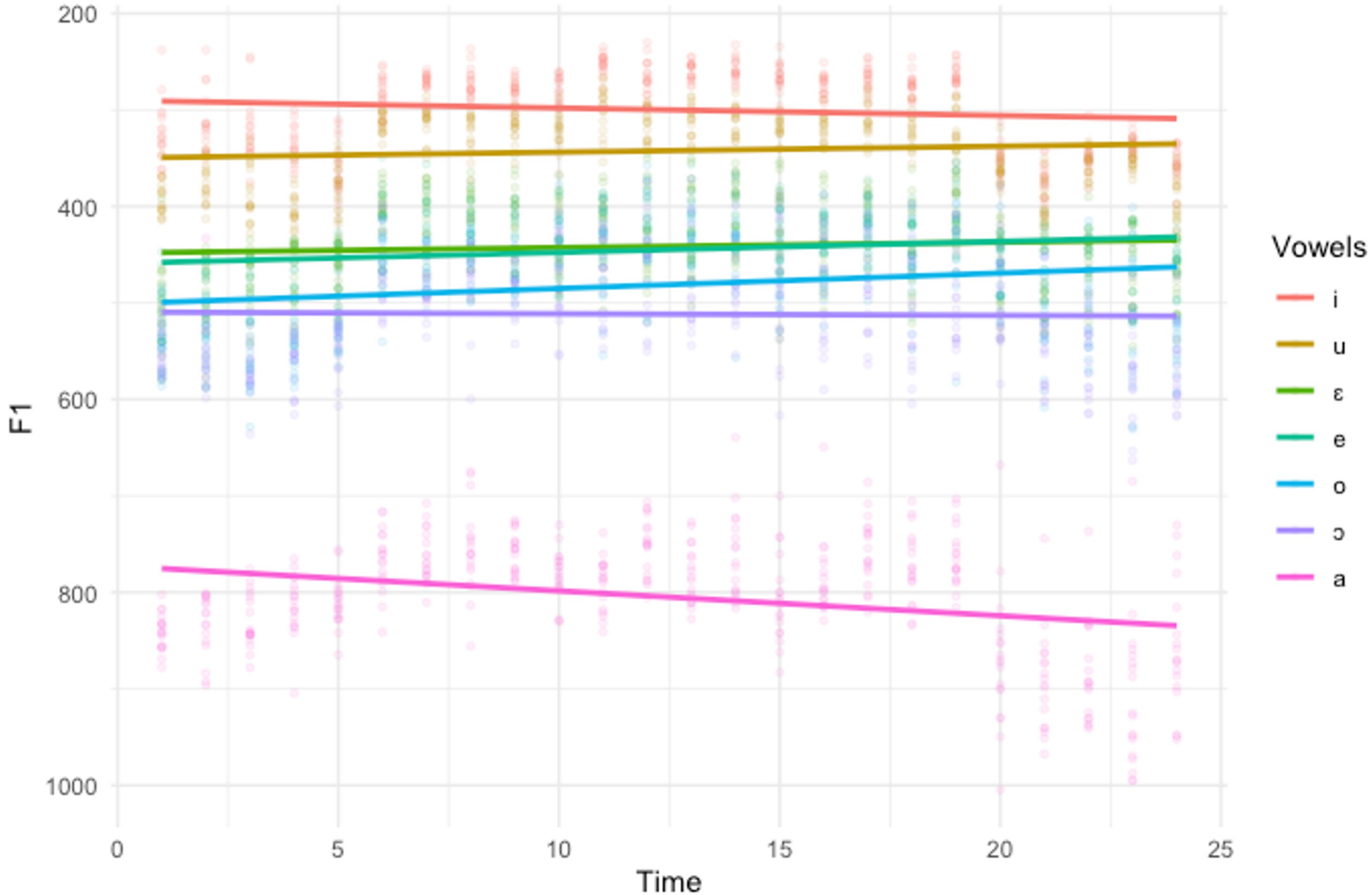
Vowel Plottings: PhonR Package.

Results

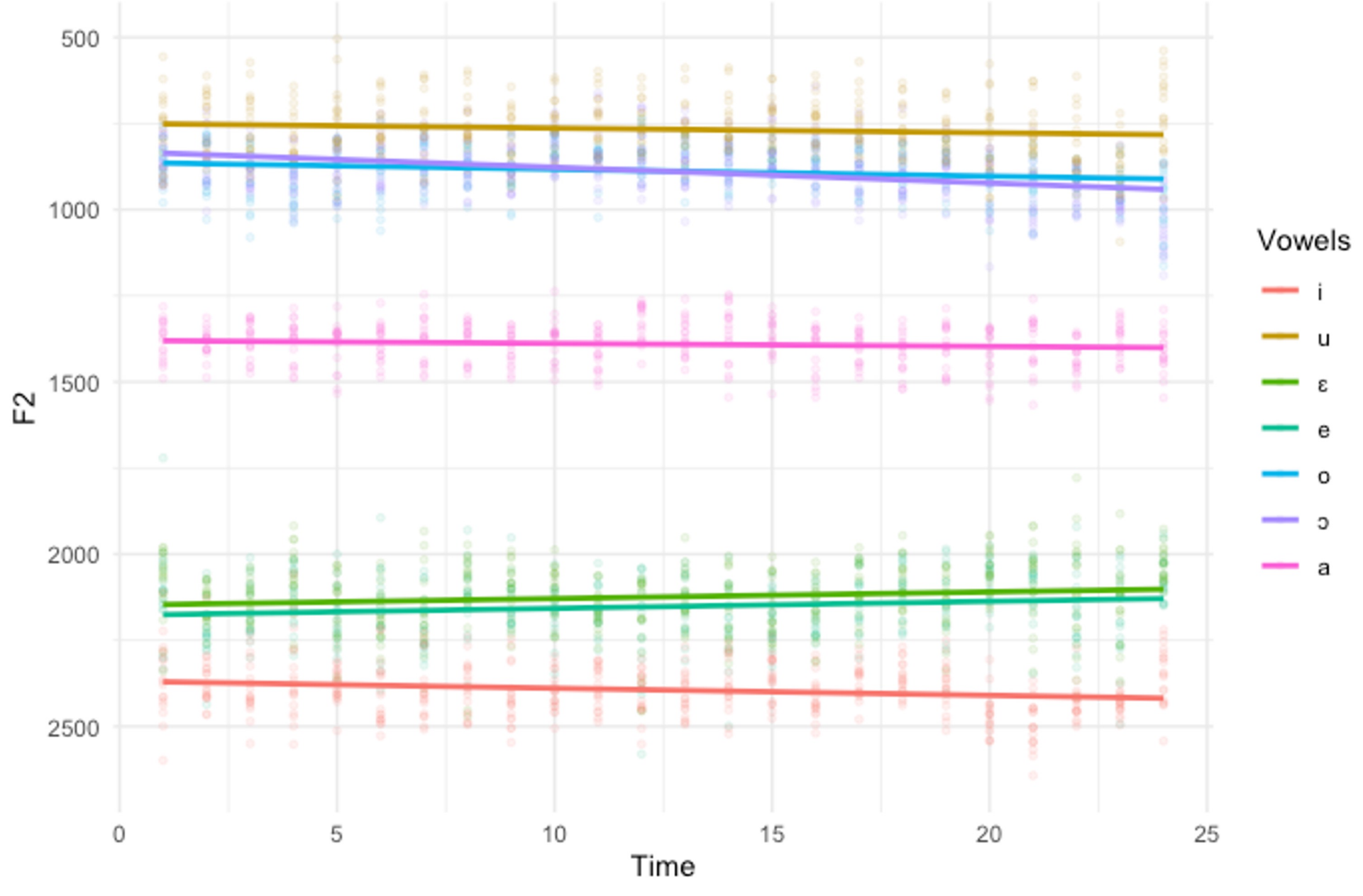

F2
3000
2500
2000
1500
1000
500
200
300
400
500
600
700
800
900
1000
1100
F1
a
e
ɛ
i
o
ɔ
u


F2
3000
2500
2000
1500
1000
500
F1
200
300
400
500
600
700
800
900
1000
1100
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

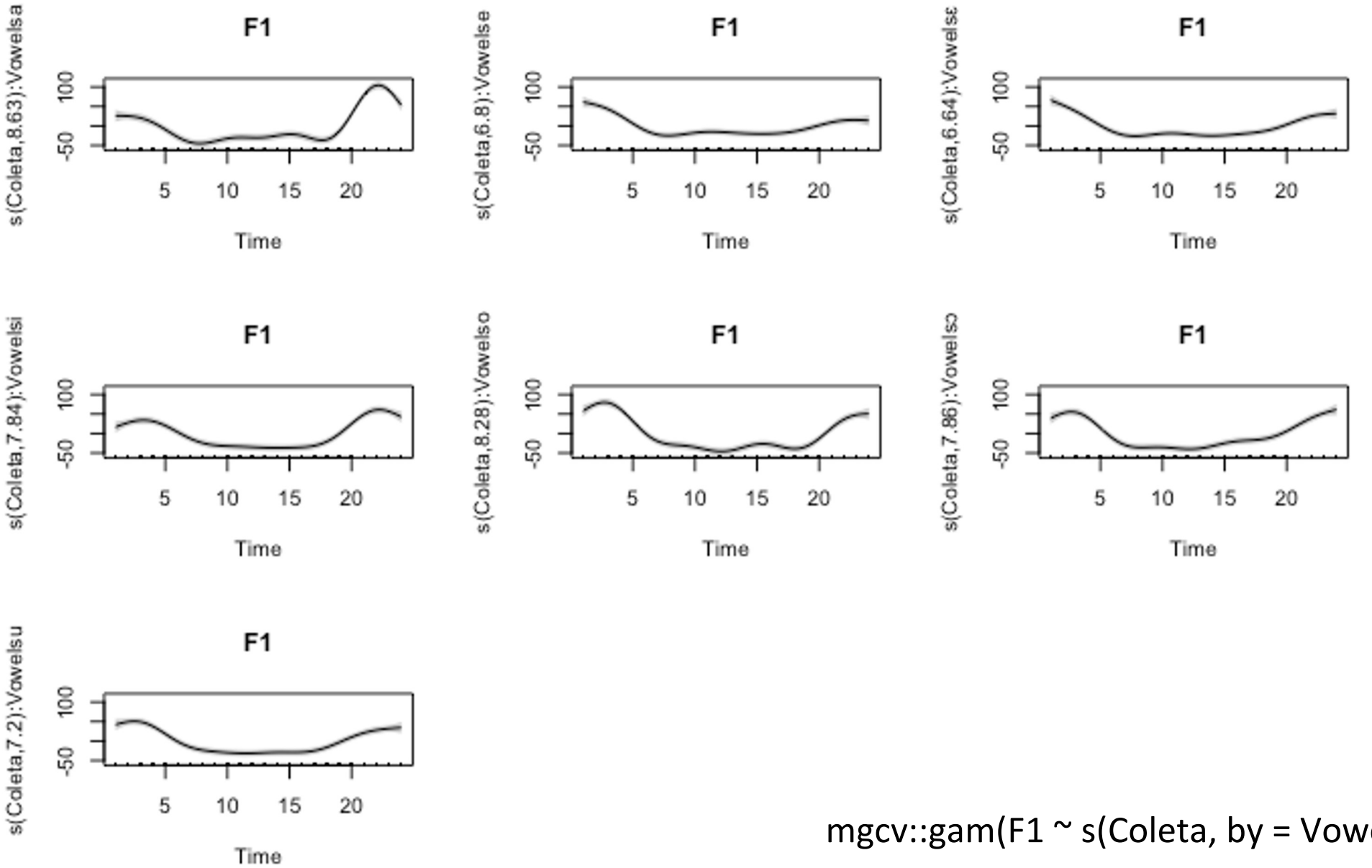
Simple Linear Model 😔



Simple Linear Model 😔

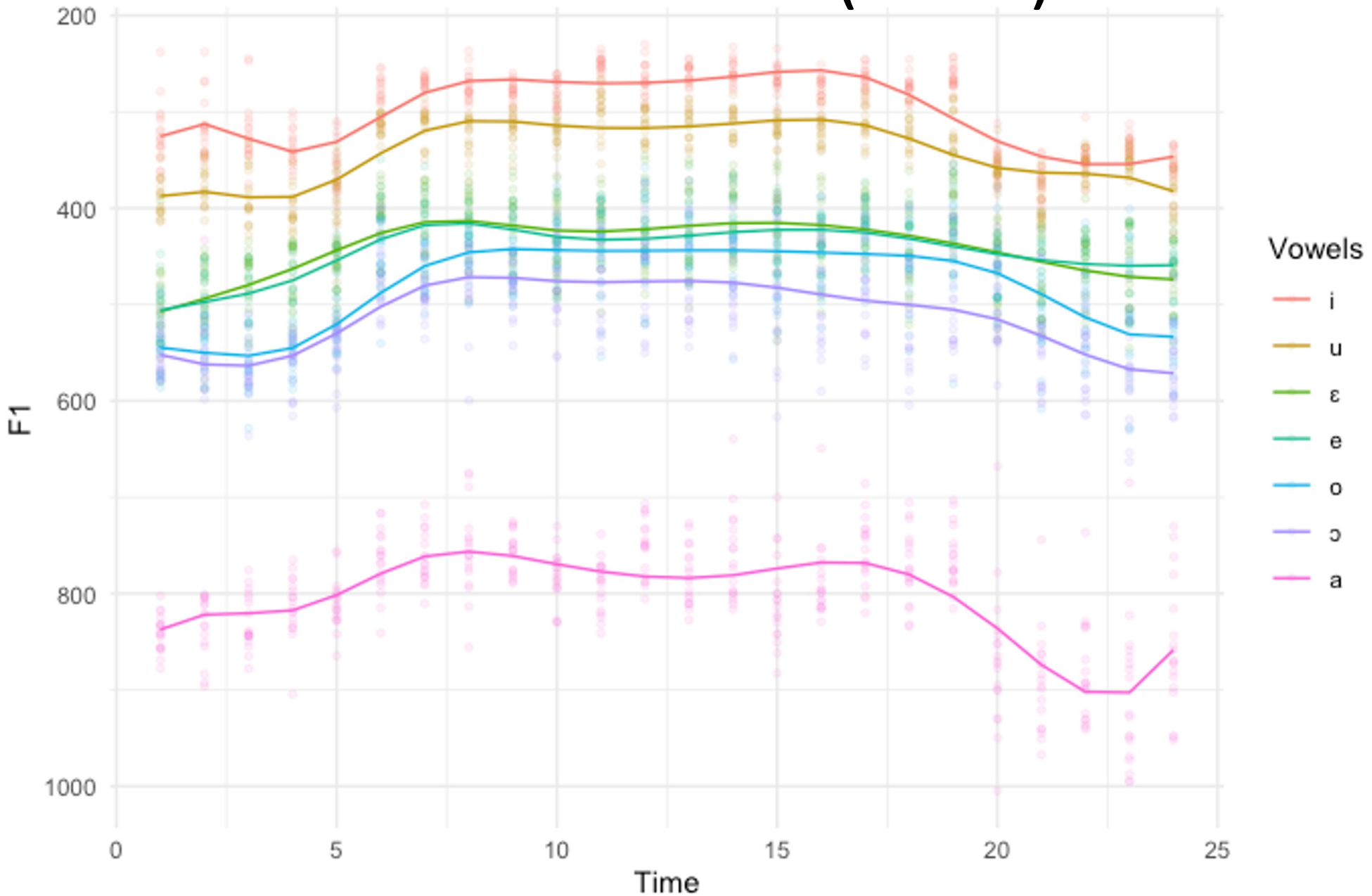


Generalized Additive Model (GAM) 🧐

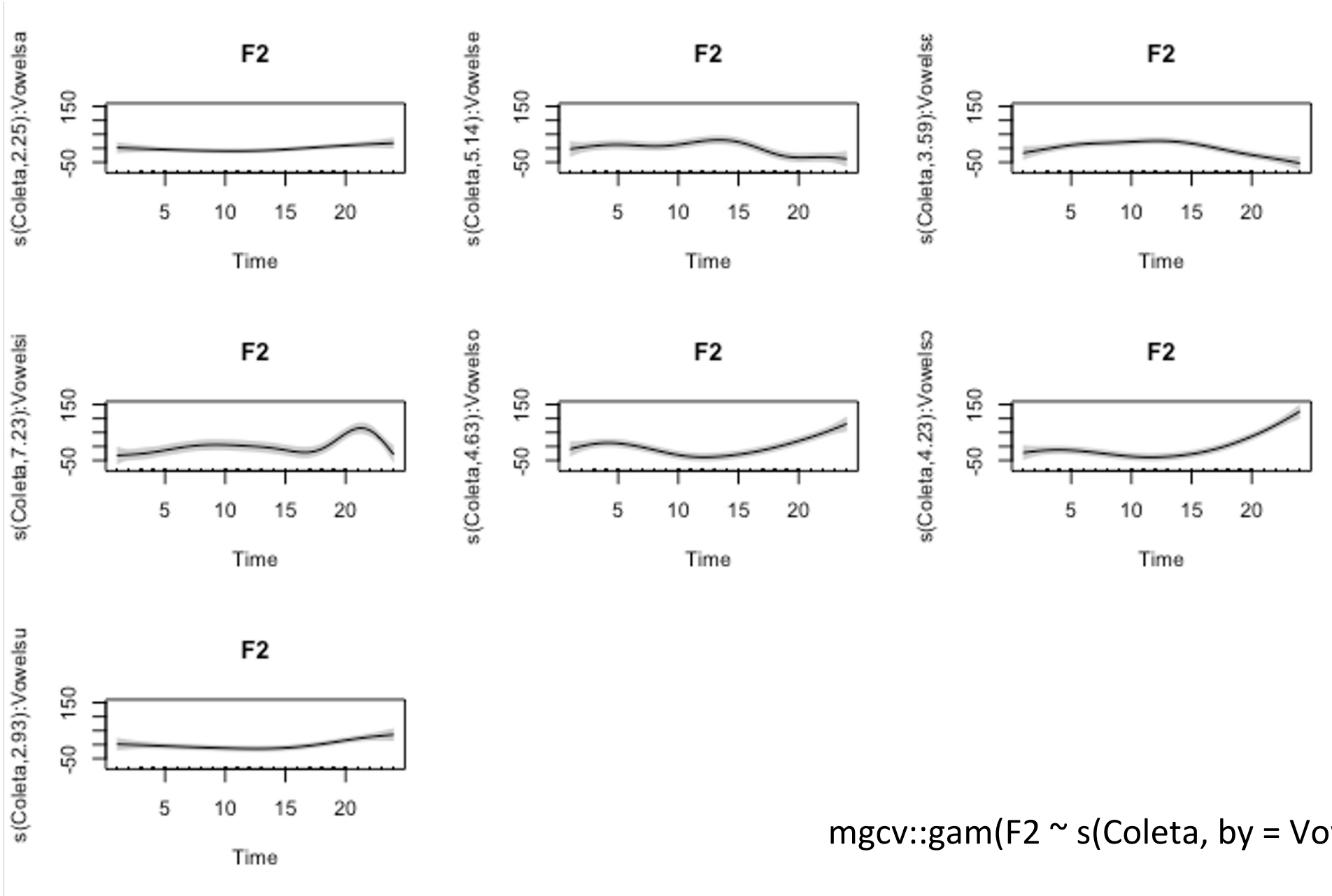


mgcv::gam(F1 ~ s(Coleta, by = Vowels) + Vowels

Generalized Additive Model (GAM) 🧐

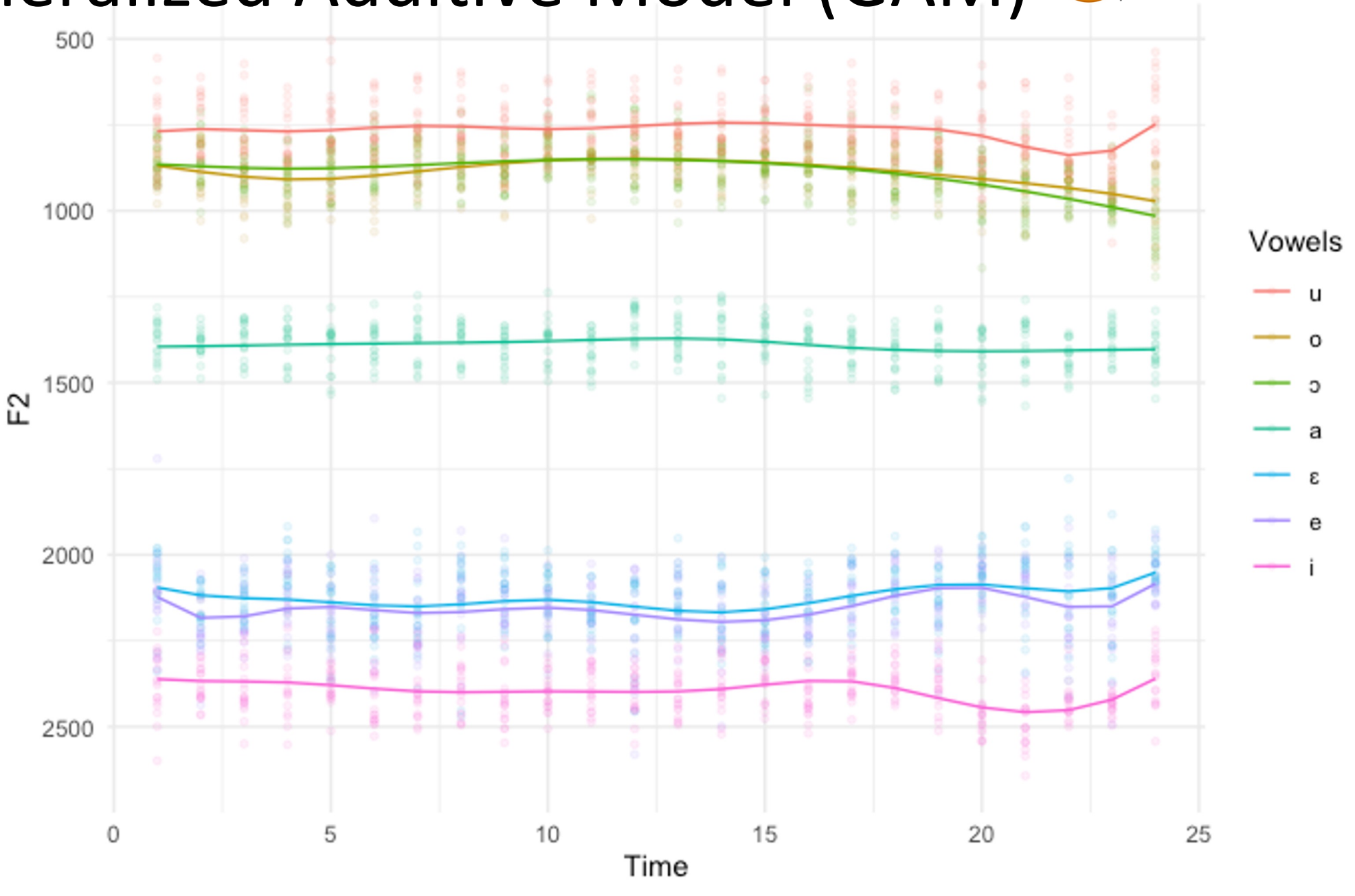


Generalized Additive Model (GAM)



mgcv::gam(F2 ~ s(Coleta, by = Vowels) + Vowels

Generalized Additive Model (GAM) 🧐



Bayesian Generalized Additive Model (GAM) 😇

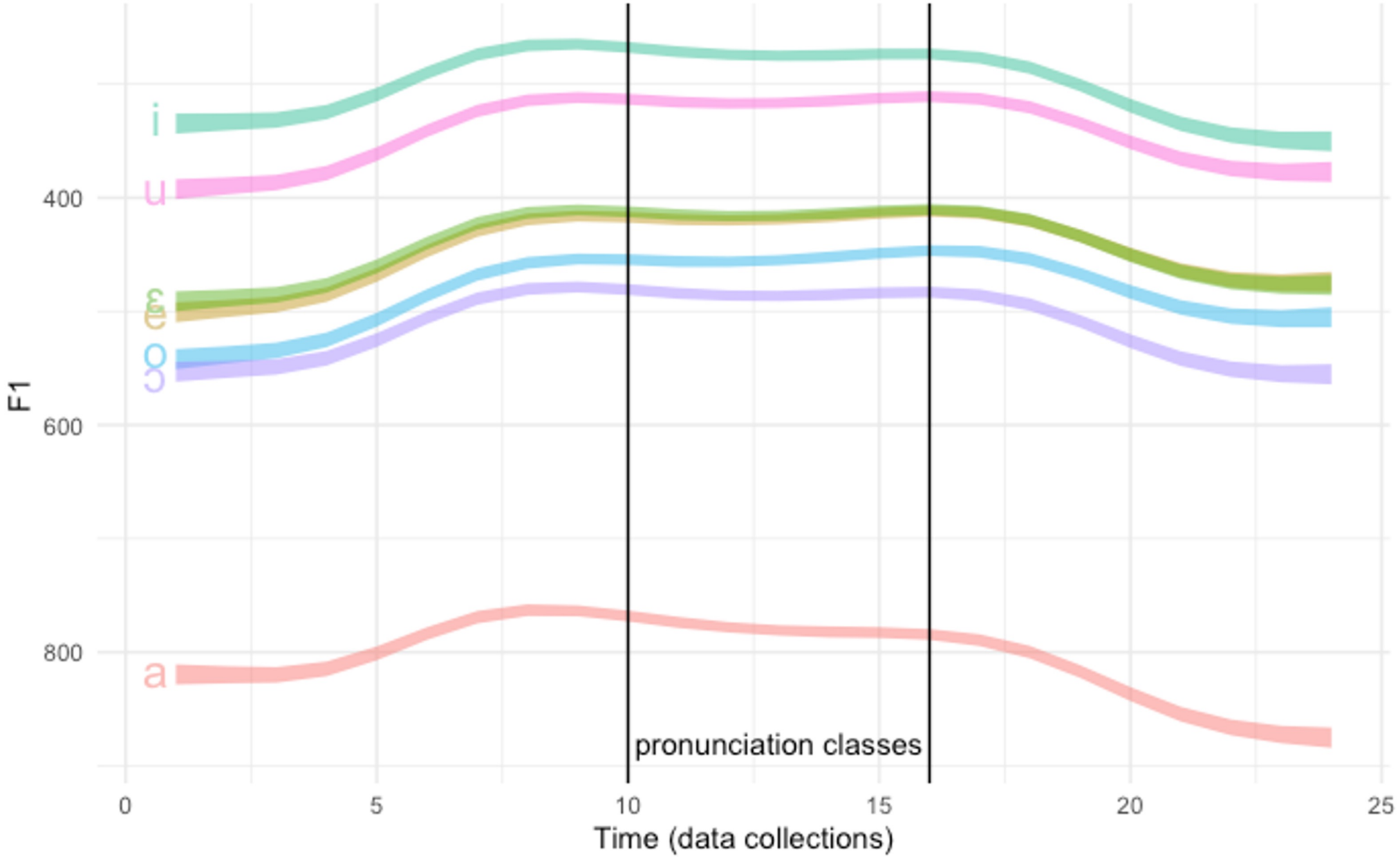
Models the probability of the parameters given the data, and not the probability of the data given a null hypothesis (no p-value);

Provides probability distributions of the coefficients instead of point estimates;

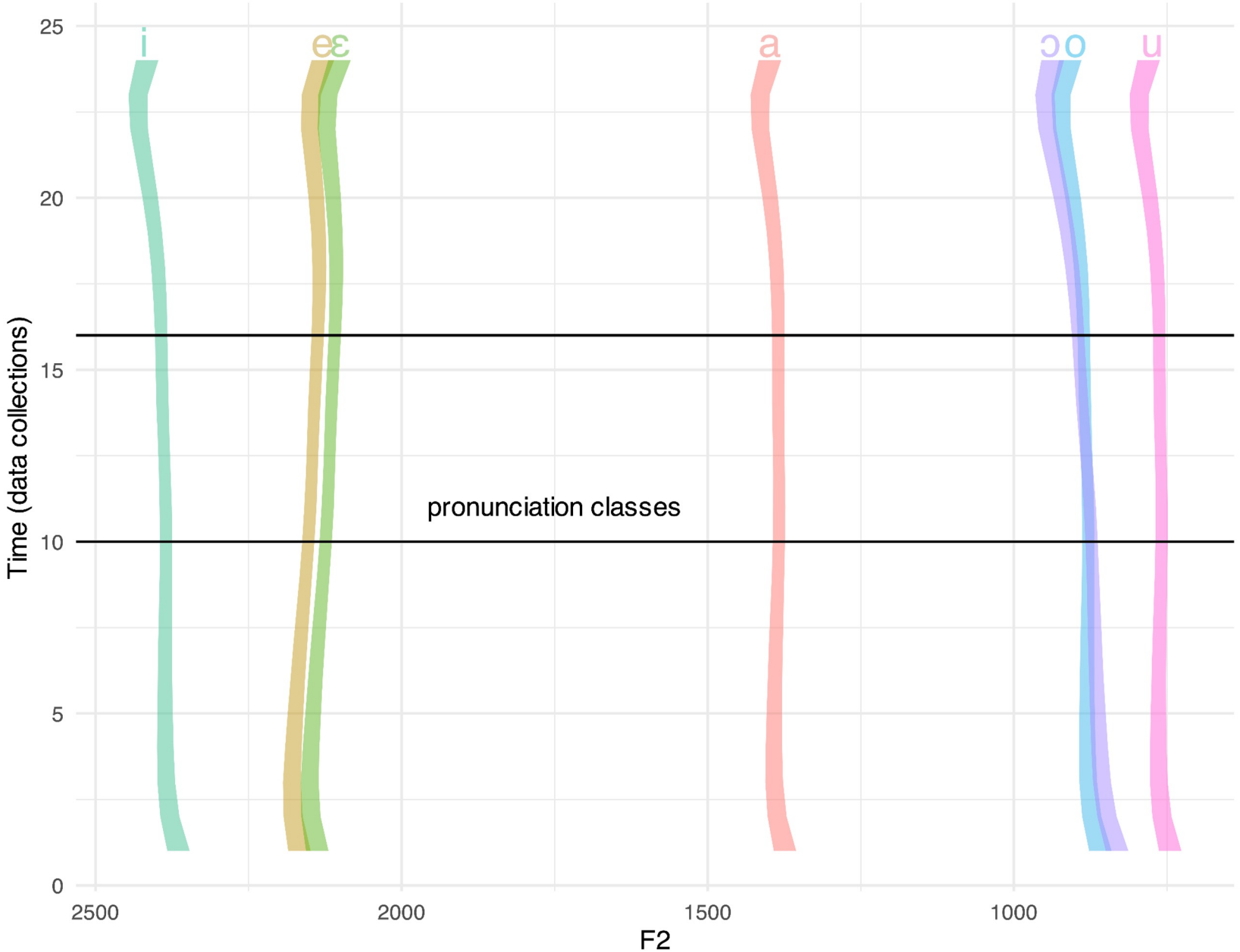
Provides credible intervals, which are more intuitive than confidence intervals;

Allows researcher to include prior knowledge of probable distributions of the data.

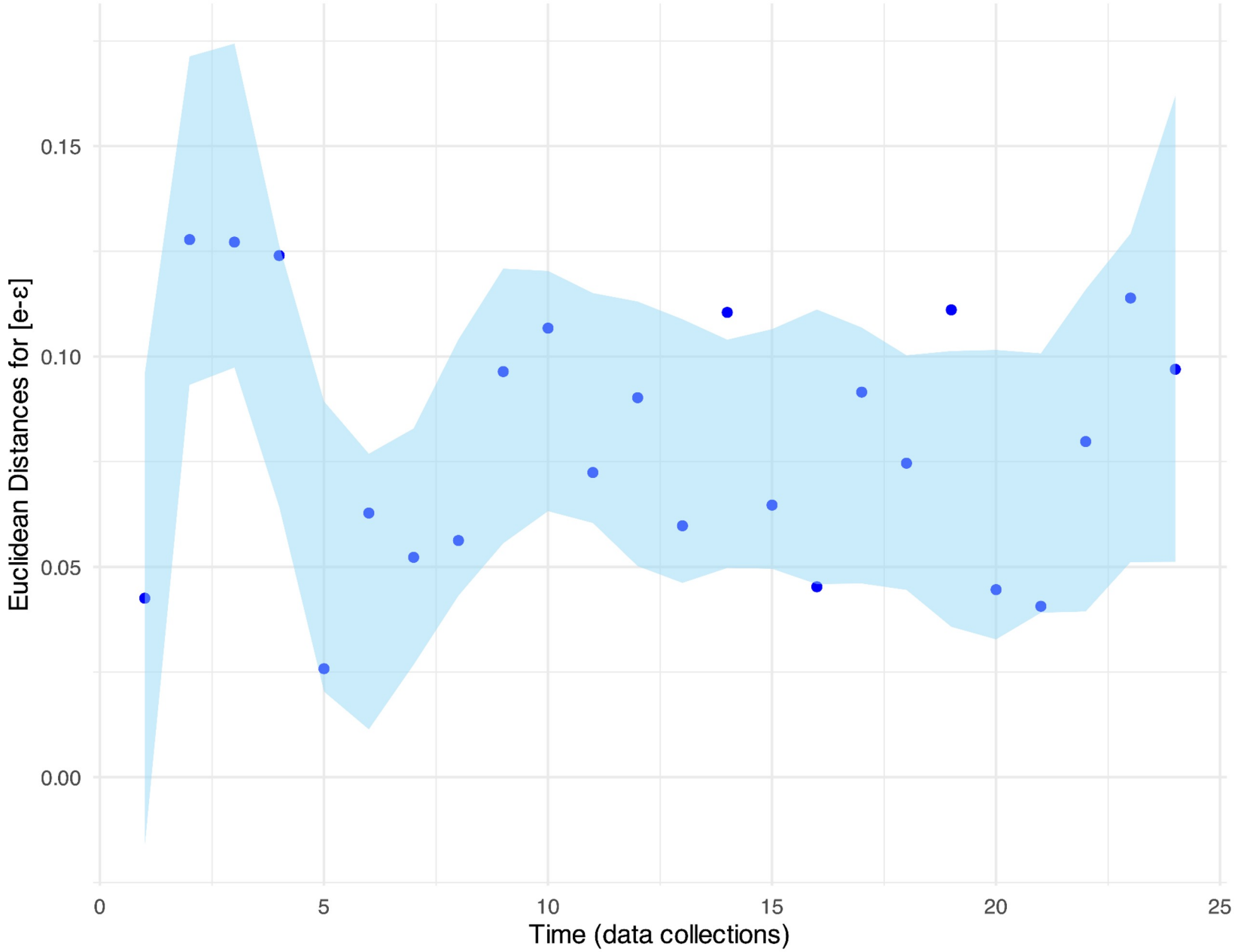
Bayesian Generalized Additive Model (GAM) 😇



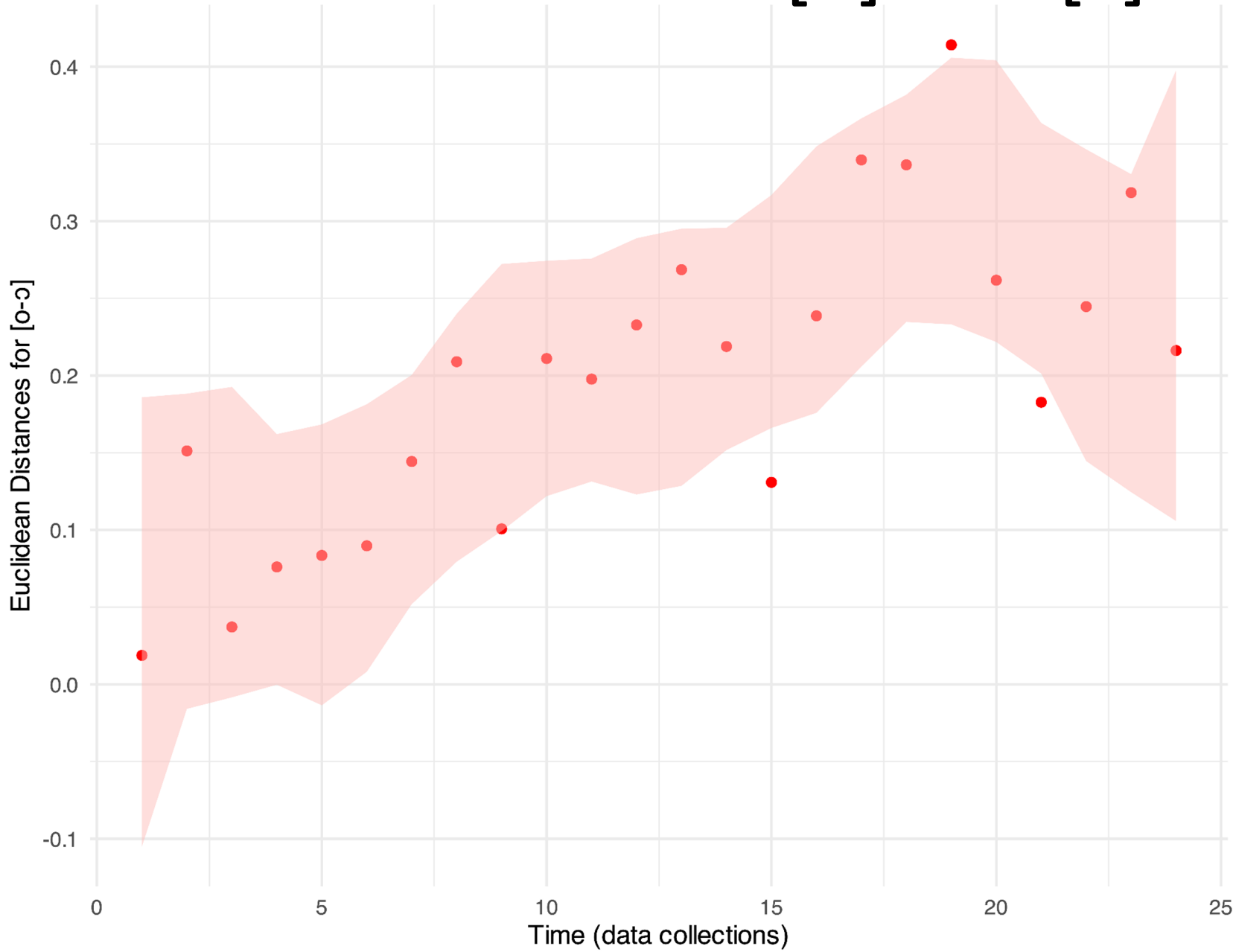
Bayesian Generalized Additive Model (GAM) 😇



Euclidean Distances between [e] and [ɛ]



Euclidean Distances between [o] and [ɔ]



Discussion

In an attempt to form new vowel categories for /ɛ/ and /ɔ/, all vowels in the learner's system interact with one another; Bayesian GAMS are able to capture these interactions.

Advantages of GAMs:

- all data points used
- all vowels in the same model
- predict curved lines

Advantages of Bayesian GAMs:

- probability of parameters given the data
- probability densities instead of point estimates
- prior knowledge may be included

References

Alves, U. K.; Brisolara, L. B.; Perozzo, R. V. (2017). Curtindo os sons do Brasil: fonética do Português do Brasil para hispanofalantes. Lisboa: LIDEL Editorial.

Boersma, P.; Weenink, D. (2019) Praat: Doing phonetics by computer – Version 6.1.16. https://www.fon.hum.uva.nl/praat/.

Pereyron, L. (2017). A produção vocálica por falantes de Espanhol (L1), Inglês (L2) e Português (L3): uma perspectiva dinâmica na (multi) direcionalidade da transferência linguística. Unpublished Doctoral Dissertation. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2017.

Taylor, W. A. (2000). Change-Point Analysis: a powerful new tool for detecting changes. Retrieved from https://variation.com/change-point-analysis-a-powerful-new-tool-for-detecting-changes/. Access on September 6, 2021.

Van DIJK, M.; Verspoor, M.; Lowie, W. (2011). Variability and DST. *In:* VERSPOOR, M.; de BOT, K.; LOWIE, W. (Eds.), *A Dynamic Approach to Second Language Development*: methods and techniques, p. 55-84. Amsterdam: John Benjamins Publishing Company.

Wood, S (2017). Generalized Additive Models: An Introduction with R, 2 edition. Chapman and Hall/CRC.

Thank you!

ronaldojr@letras.ufc.br

ukalves@gmail.com