



Accelerating the Analysis of your Audio Recordings with Untrained Forced Speech Alignment

Rolando Coto-Solano¹, Sally Akevai Nicholas², Samantha Wray³, Tyler Peterson⁴

¹ Te Whare Wānanga o te Ūpoko o te Ika a Māui / Victoria University of Wellington

² Te Kunenga ki Pūrehuroa / Massey University

³ New York University Abu Dhabi

⁴ Arizona State University

6th International Conference on Language Documentation & Conservation (ICLDC)

Connecting Communities, Languages & Technology

University of Hawaii at Manoa, February 28-29 2019

Workshop Roadmap

- 1) Introduction: What is forced alignment?
- 2) What will we produce here today?
- 3) What are the necessary files you need to make as input?

Workshop Roadmap

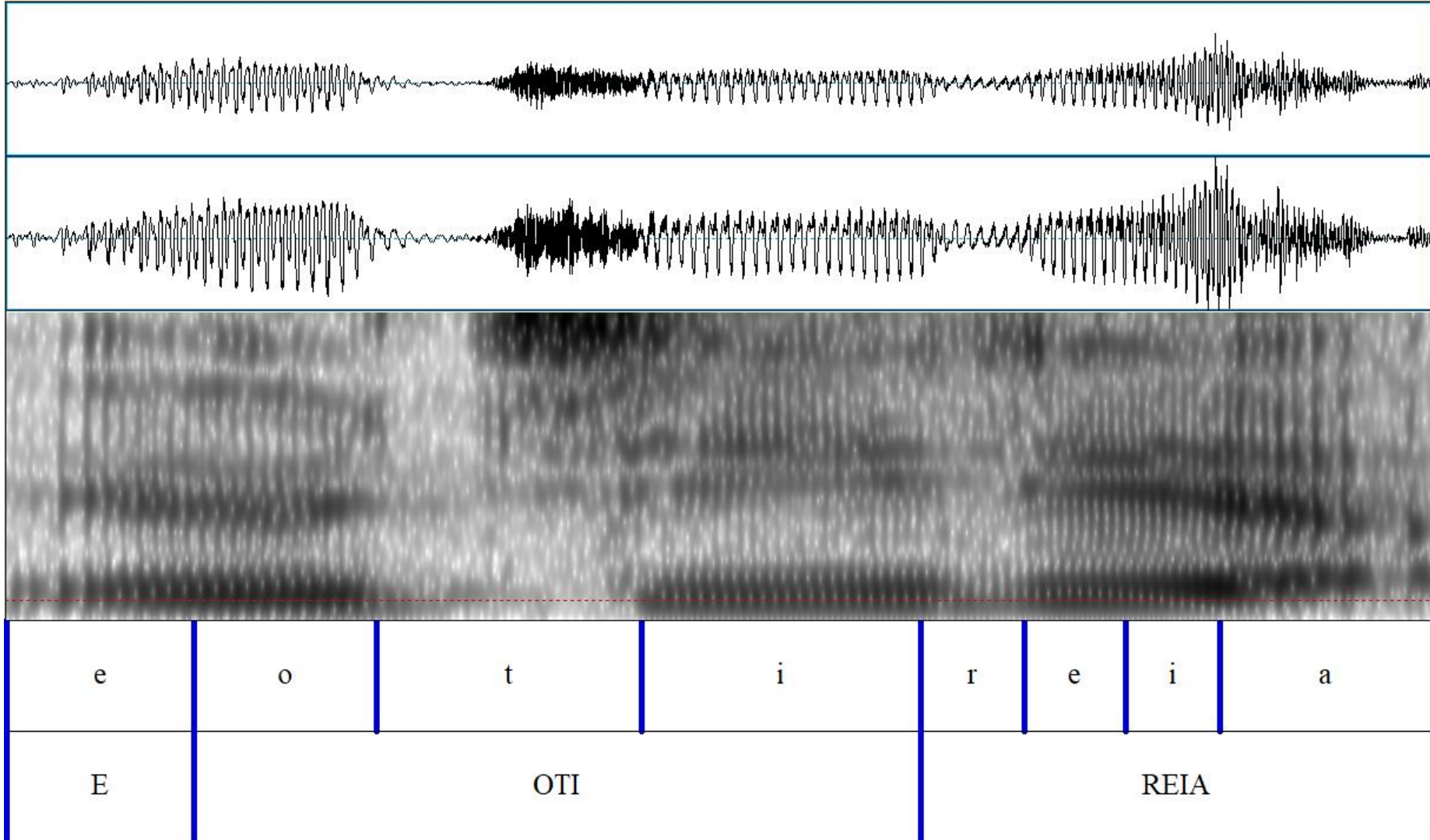
- 1) Introduction: What is forced alignment?
- 2) What will we produce here today?
- 3) What are the necessary files you need to make as input?

We start with an audio recording and a transcription

Jean	CIM	0.14	1.0	e oti reia
Jean	CIM	1.6	2.6	ax
Jean	CIM	4	4.96	te openga i reia o texrax



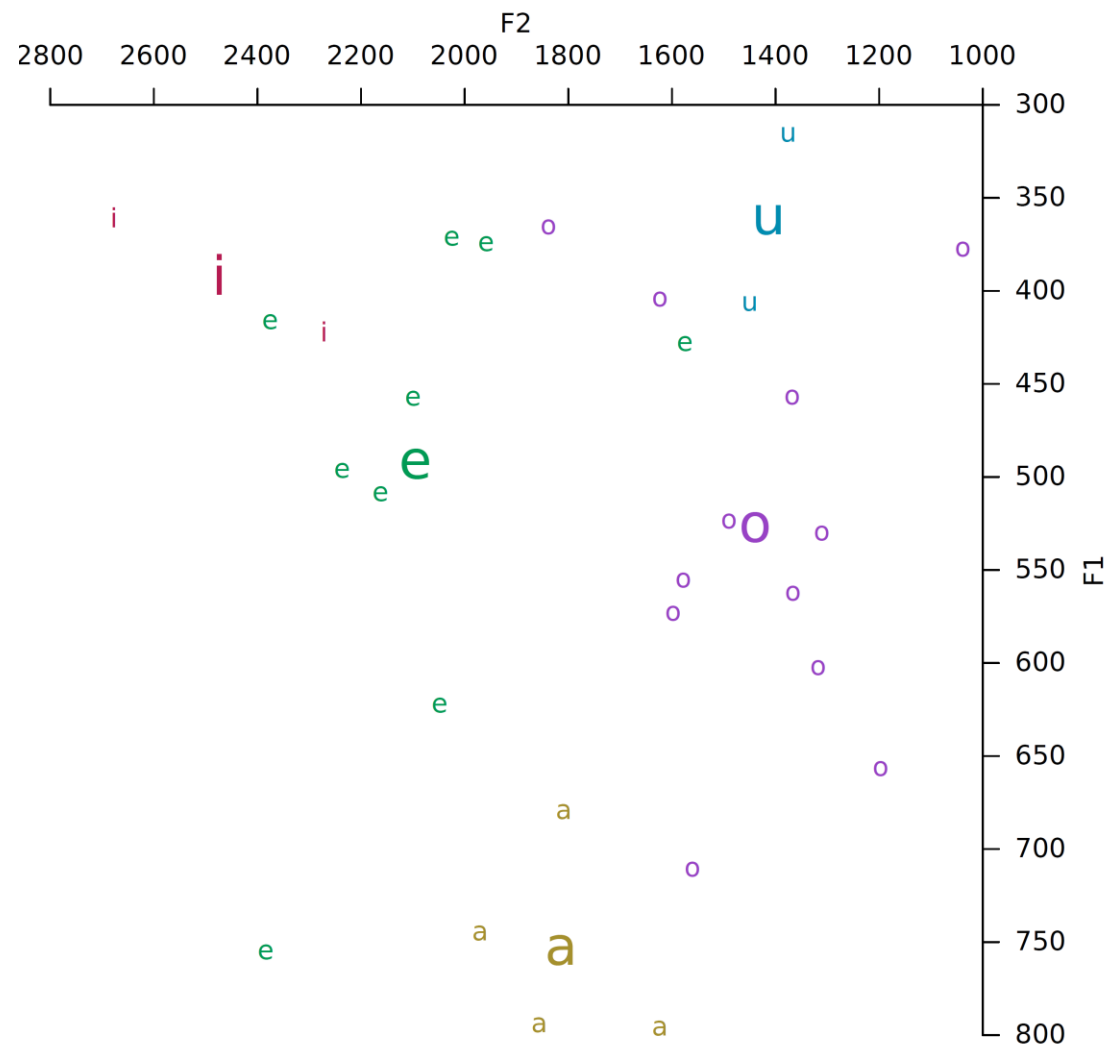
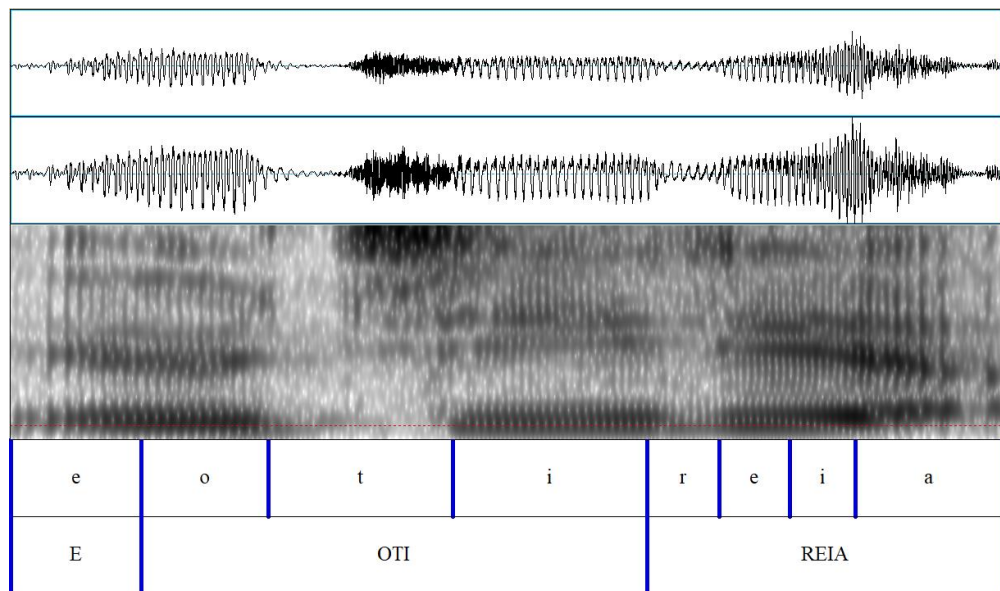
We will use the recording and the transcription to create an “aligned transcription”



Workshop Roadmap

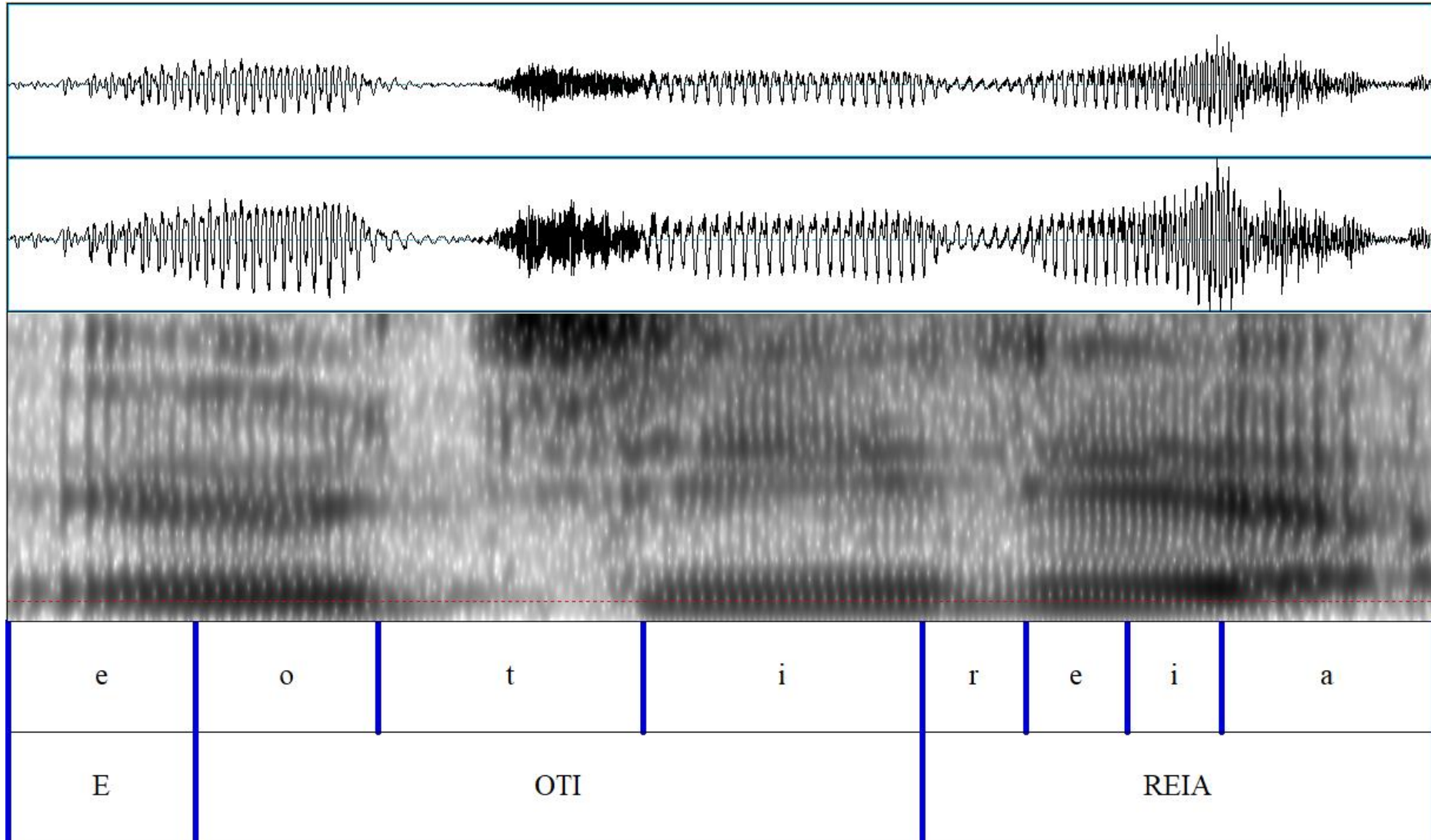
- 1) Introduction: What is forced alignment?
- 2) What will we produce here today?**
- 3) What are the necessary files you need to make as input?

Aligned transcription, phonetic information, vowel inventory

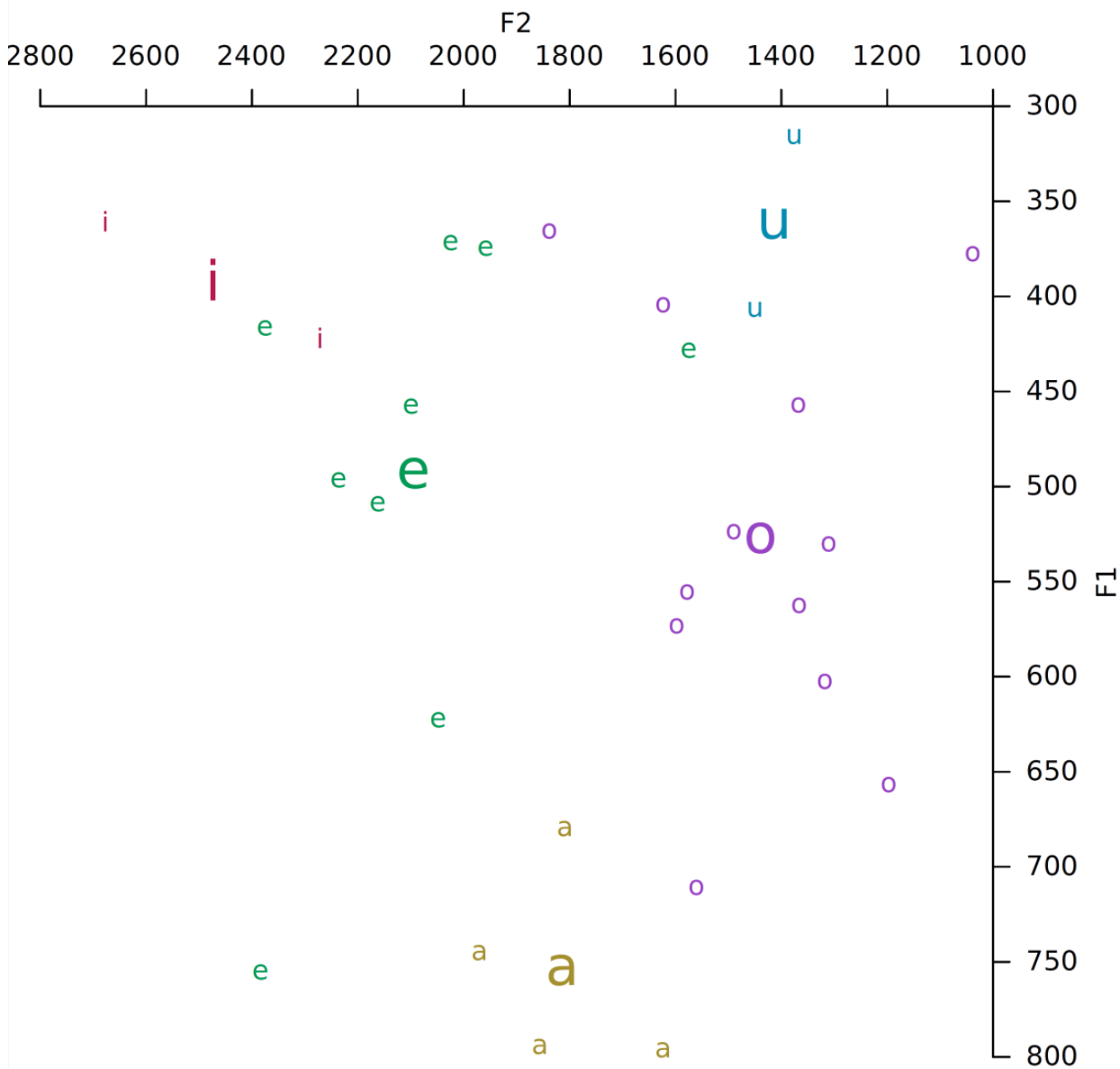


Filename	TextGridLabel	Word	PreviousLabel	FollowingLabel	start	end	duration	f0_0%point	f0_10%point	f
cim-jane	sp	sp	-	sp	0	0.153	0.153	--undefined--	--undefined--	--undefined--
cim-jane	sp	sp	sp	e	0.153	0.2591471270778328	0.10614712707783278	--undefined--	--undefined--	--und
cim-jane	e	E	sp	o	0.2591471270778328	0.32980889799997987	0.0706617709221471	--undefined--	--undefined--	--undefin
cim-jane	o	OTI	e	t	0.32980889799997987	0.3982964605860609	0.06848756258608102	167.51815148247317	168.59659	
cim-jane	t	OTI	o	i	0.3982964605860609	0.4983100440450998	0.10001358345903893	175.38913916490898	172.78360	
cim-jane	i	OTI	t	r	0.4983100440450998	0.603	0.10468995595490016	218.090294315789	216.5089086934359	2
cim-jane	r	REIA	i	e	0.603	0.6422463857190984	0.03924638571909844	228.44061768327037	226.5568075331704	

We will use the recording and the transcription to create an “aligned transcription”

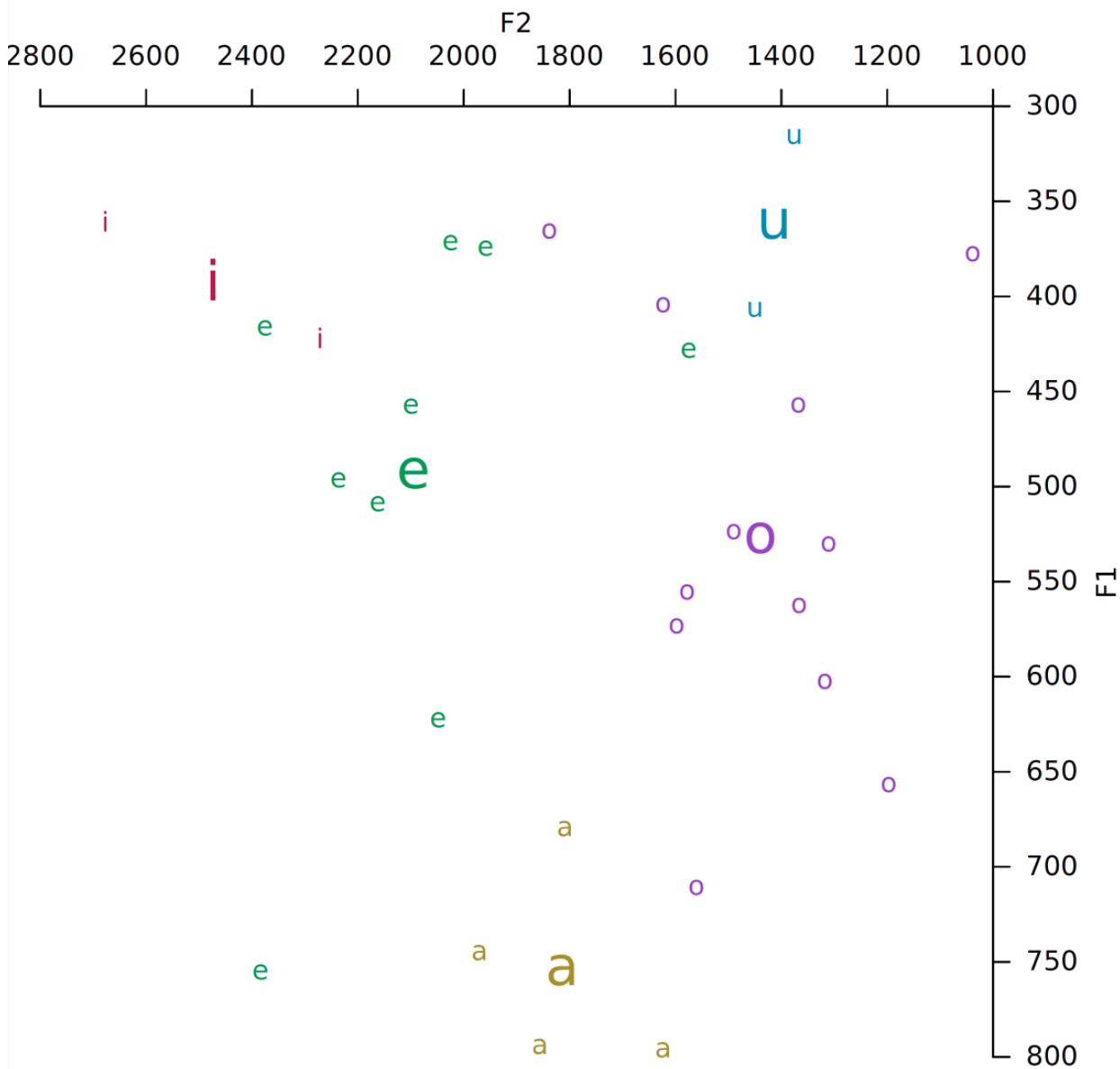


We will ultimately produce a “vowel triangle”



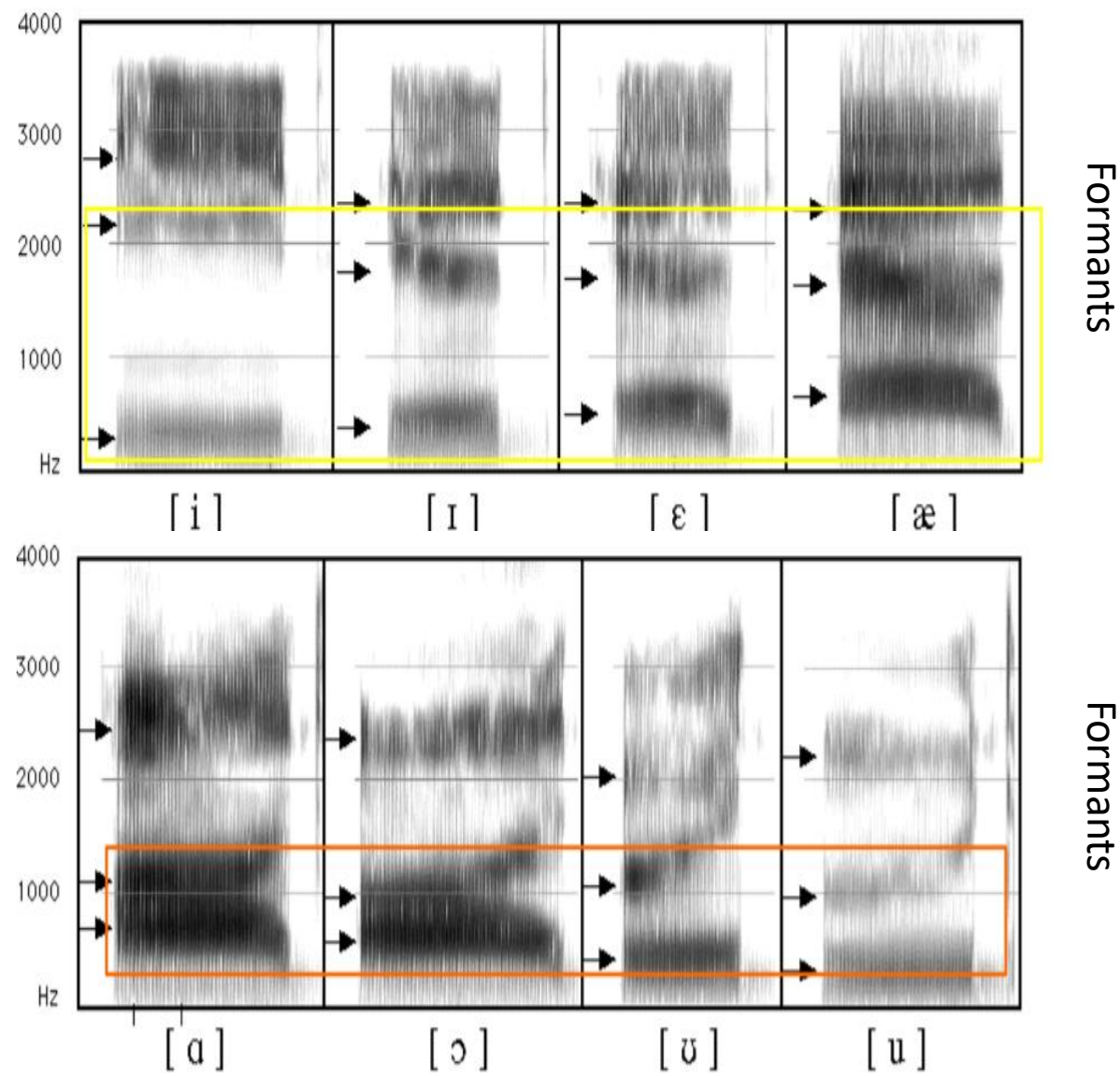
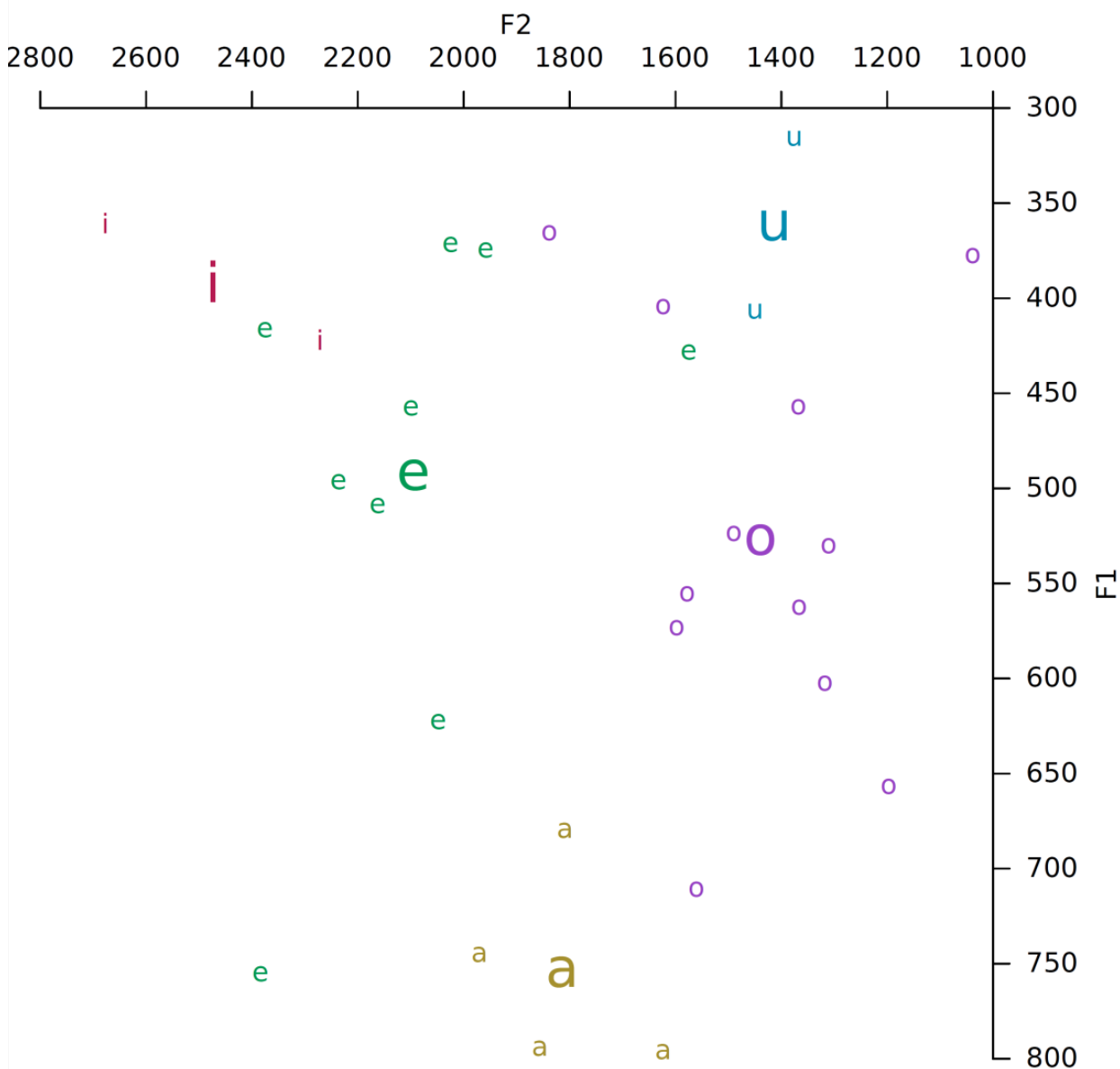
Automatically plotting
the variation of vowels
in your language!

We will ultimately produce a “vowel triangle”

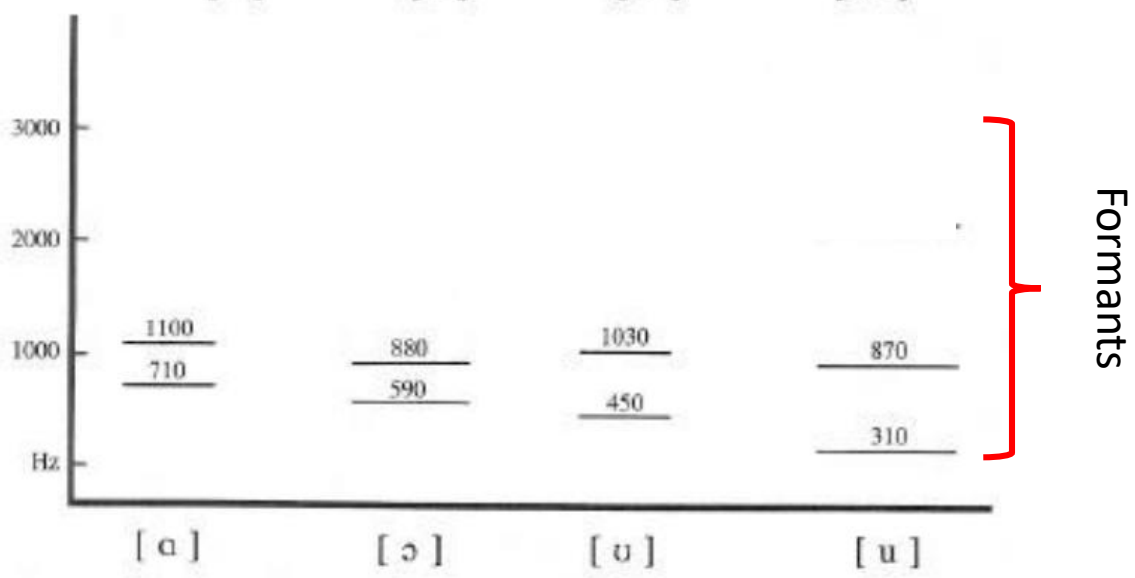
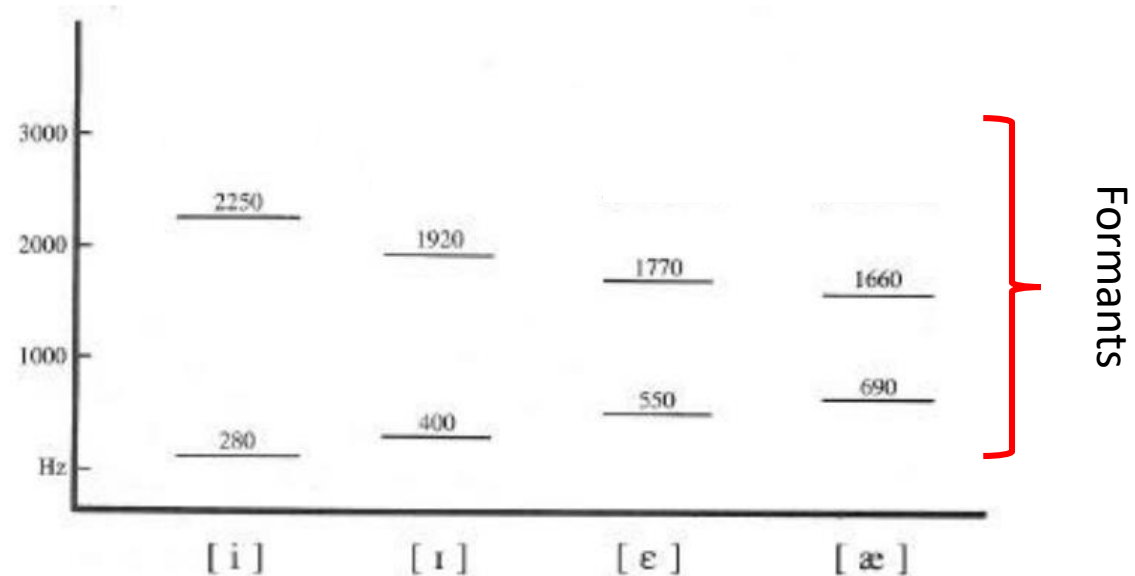
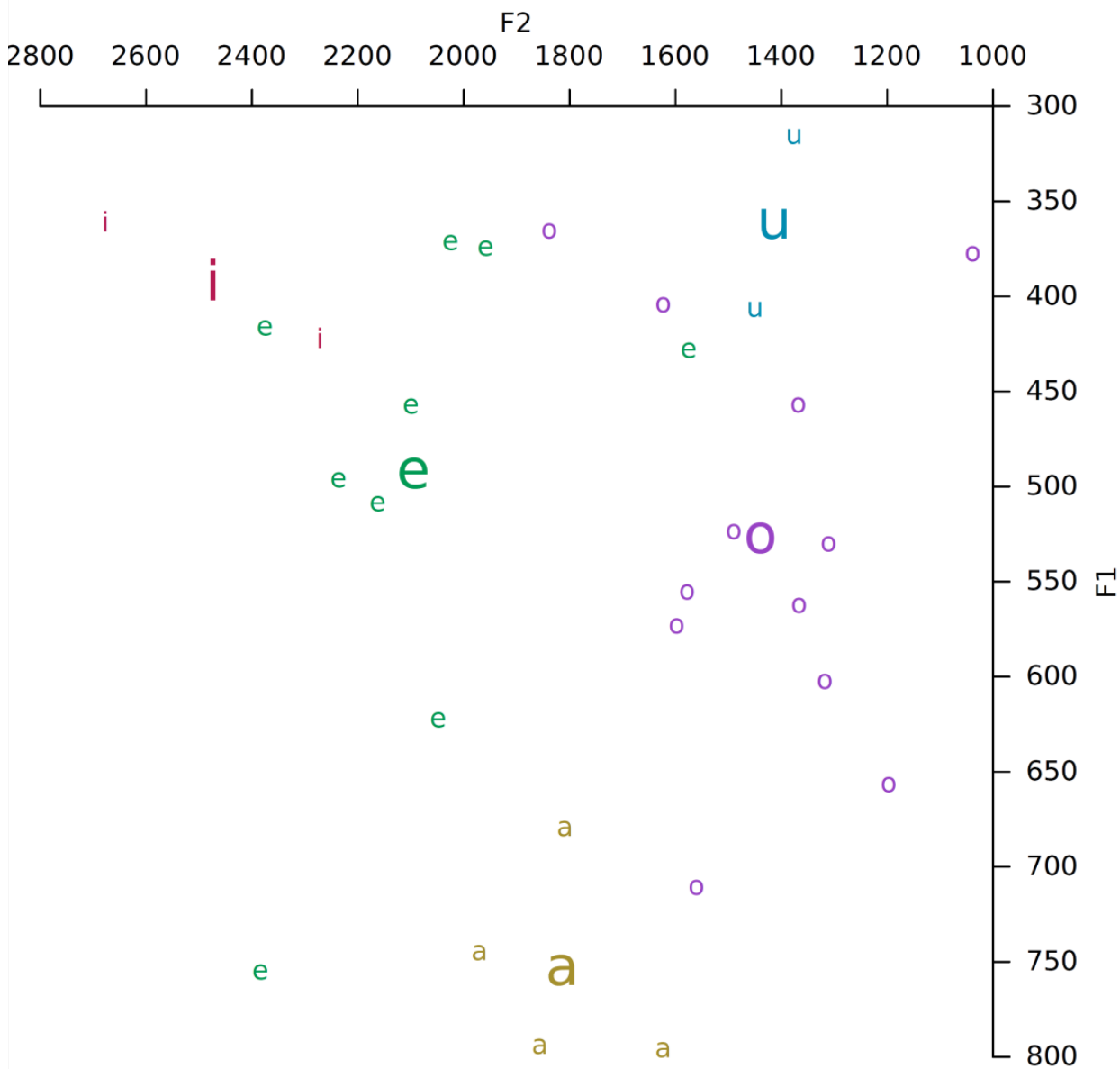


In speech, the resonant frequencies of the vocal tract are called **formants**

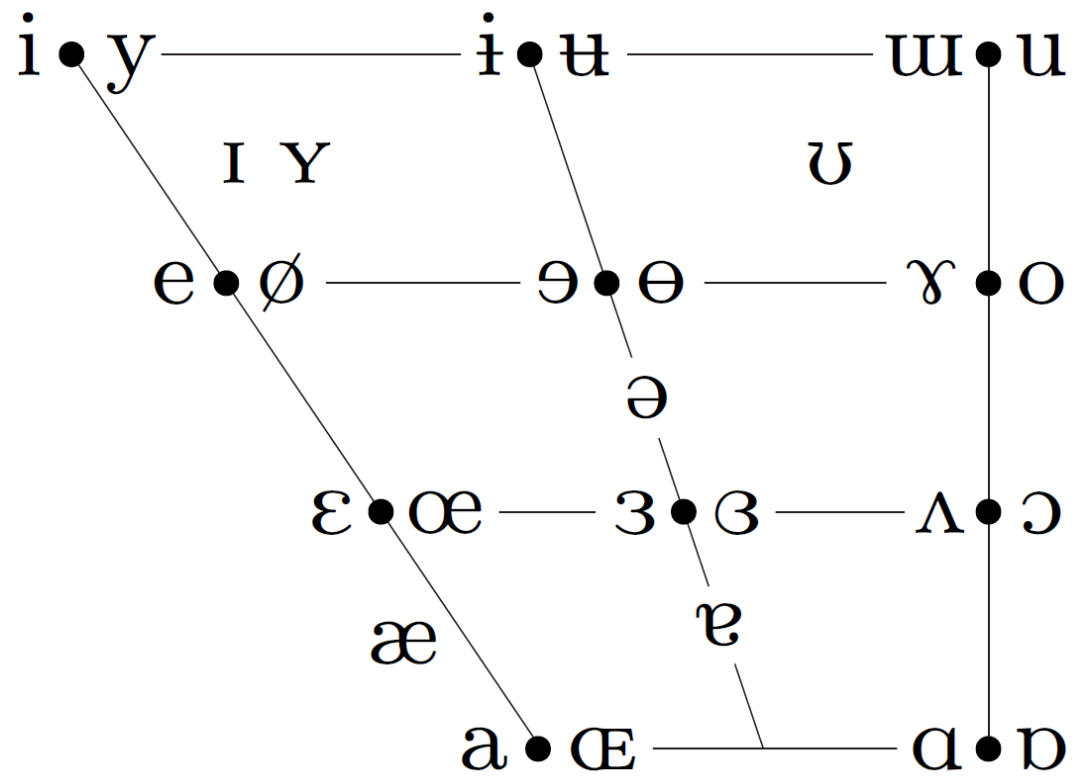
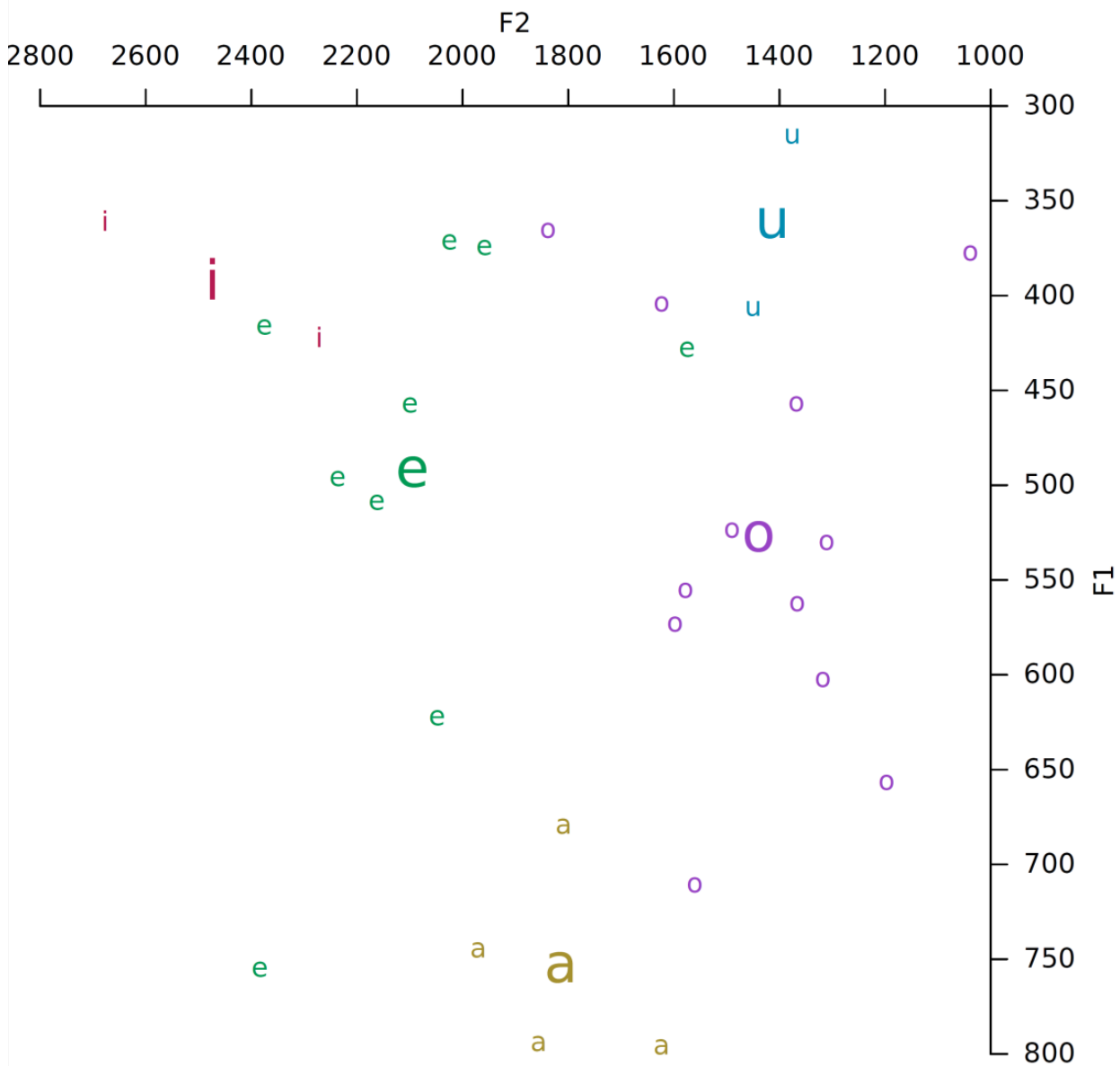
We will ultimately produce a “vowel triangle”



We will ultimately produce a “vowel triangle”



We will ultimately produce a “vowel triangle”



Workshop Roadmap

- 1) Introduction: What is forced alignment?
- 2) What will we produce here today?
- 3) What are the necessary files you need to make as input?

We'll make a short recording of speech



First we need to extract (or create) a recording in your language.

It should be about 5 seconds long.

Using Praat, premier software for phonetic analysis

We'll make the transcription

Tab



Tab



Tab



Tab



Jean	CIM	0.14	1.0	e oti reia
Jean	CIM	1.6	2.6	ax
Jean	CIM	4	4.96	te openga i reia o texrax



Speaker
name



"Tier"
name



Beginning
of phrase



End of
phrase



Words in
the phrase

Transcribing using the ARPAbet

ARPAbet: Set of phonetic transcription codes

Extremely basic; for the sake of the training algorithm, we may have to simplify the sound system of your language

te openga i reia o tērā

te openga i reia o texrax

T EH1 OW1 P EH1 NG AE1 IY1 R EH1 IY1 AE1 OW1 T EH1 R AE1

Transcribing using the ARPAbet

ARPAbet: Set of phonetic transcription codes

Extremely basic; for the sake of the training algorithm, we may have to simplify the sound system of your language

te openga i reia o tērā

te openga i reia o texrax

T EH1 OW1 P EH1 NG AE1 IY1 R EH1 IY1 AE1 OW1 T EH1 R AE1

Stress markers accompany vowels; for best performance we'll mark everything as primary stress

Transcribing using the ARPAbet

ARPAbet: Set of phonetic transcription codes

Extremely basic; for the sake of the training algorithm, we may have to simplify the sound system of your language

te openga i reia o tērā

te openga i reia o texrax

T EH1 OW1 P EH1 NG AE1 IY1 R EH1 IY1 AE1 OW1 T EH1 R AE1

The ARPAbet is an imperfect system biased towards Standard American English. **R** here means the SAE /ɹ/ not /r/. However, the alignment algorithm will prevail!

We need to make the dictionary

