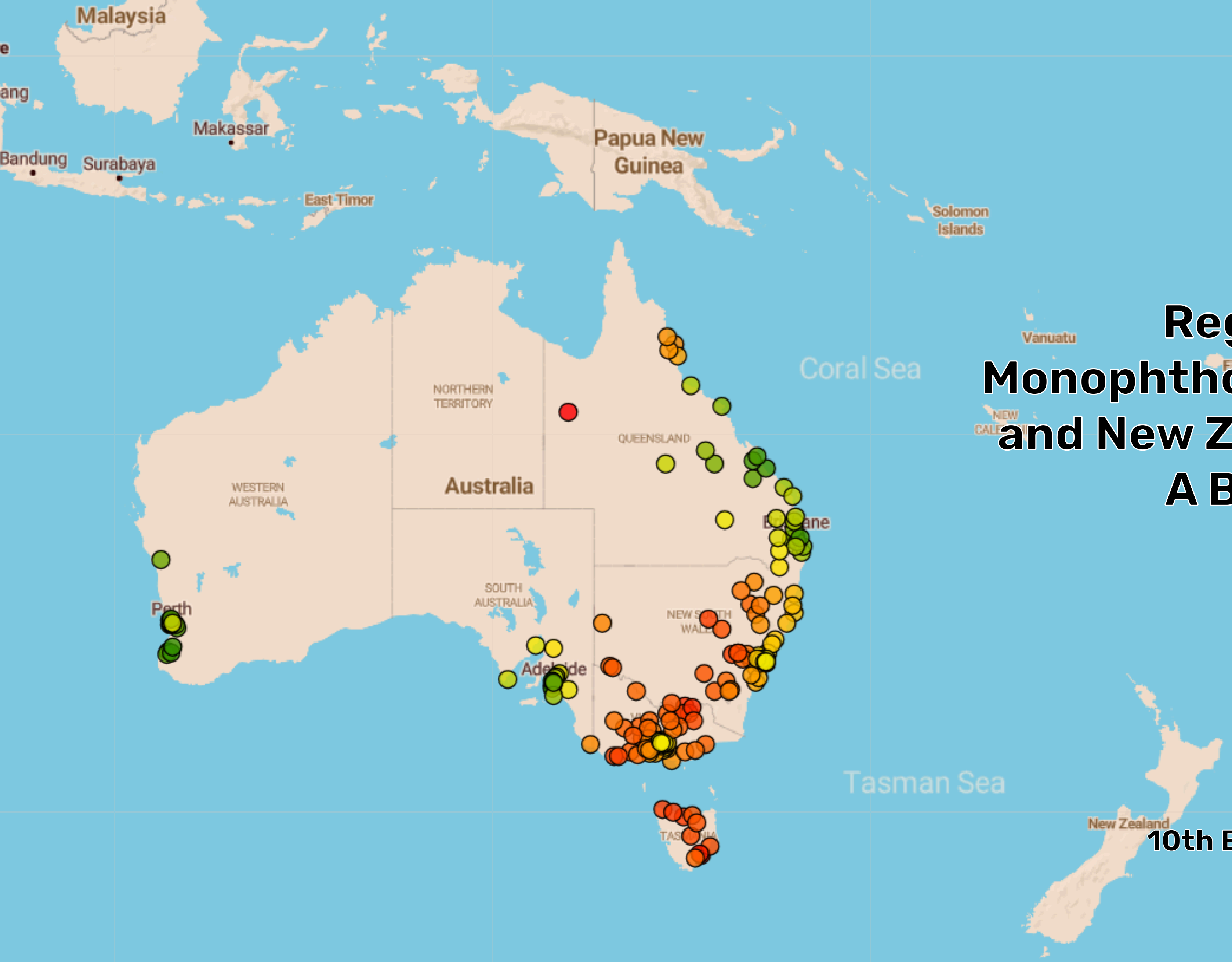


Regional Variation in Monophthongs in Australian and New Zealand Englishes: A Big Data Approach

Steven Coats
University of Oulu, Finland
steven.coats@oulu.fi
10th BICLCE Conference, Alicante
September 27th, 2024

Outline

1. Background: Regional variation in Australian and New Zealand Englishes
2. Data: CoANZSE
3. Method: Scripting pipeline for collection of data, alignment, formant extraction, and spatial analysis
4. Preliminary results
5. Caveats, summary

Slides for the presentation are on my homepage at https://cc.oulu.fi/~scoats

Background

Traditional view: Regional variation is limited in AUS and NZ. "Australia is, generally speaking, linguistically unified" (Mitchell & Delbridge 1965: 13).

Australia

- Lexical items, e.g. *potato cake/scallop/fritter* (Bryant 1989)
- Realization of words like *dance* ([dæns] or [dans]) can differ, /a/ more common in Adelaide than Melbourne, Sydney, and Brisbane (Horvath & Horvath 2001)

New Zealand

- Lexical divergence of children's playground vocabulary items (Bauer & Bauer 2002)
- Rhoticity in the South Island (Kennedy 2006; Marsden 2013)

Recently: Some regional phonetic variation exists in Australia

- Analysis of 5,722 vowel tokens in hVd words by 109 younger speakers from Melbourne, Sydney, Adelaide, and Perth (Cox & Palethorpe 2019)
- Some evidence for distinctive realization of GOAT diphthong in Adelaide, other vowels with mixed results

This study: Investigation of regional variation based on audio and transcripts from YouTube videos indexed in the Corpus of Australian and New Zealand Spoken English

CoANZSE (Coats 2022, 2024)

- ASR transcripts from YouTube channels of regional and local councils
- Many recordings are meetings: advantages in terms of representativeness and comparability
- Speaker place of residence (cf. videos collected based on place-name search alone)
- Topical contents and communicative contexts comparable
- Content either in the public domain (US) or can be used under "fair use" or "fair dealings" provisions of copyright law (e.g. Australian Copyright Act of 1968)

Example video

Maranoa Regional Council - Ordinary Meeting - 24 Novemb...

WebVTT file

```
WEBVTT
Kind: captions
Language: en

00:00:01.160 --> 00:00:06.550 align:start position:0%

[Music]

00:00:06.550 --> 00:00:06.560 align:start position:0%
[Music]


00:00:06.560 --> 00:00:08.150 align:start position:0%
[Music]
uh<00:00:06.960><c> welcome</c>

00:00:08.150 --> 00:00:08.160 align:start position:0%
uh welcome


00:00:08.160 --> 00:00:10.950 align:start position:0%
uh welcome
i'd<00:00:08.320><c> like</c><00:00:08.480><c> to</c><00:00:08.639><c> open</c><00:00:08.880><c> the</c><00:00:09.040><c> meeting</c><00:00:09.360><c> at</c><00:00:09.519><

00:00:10.950 --> 00:00:10.960 align:start position:0%
i'd like to open the meeting at 9 12 a.m


00:00:10.960 --> 00:00:13.190 align:start position:0%
i'd like to open the meeting at 9 12 a.m
thank<00:00:11.200><c> you</c><00:00:11.280><c> for</c><00:00:11.440><c> your</c><00:00:11.599><c> attendance</c>
```

Data format

	country	state	name	channel_name	channel_url	video_title	video_id	upload_date	video_length	text_pos	location	latlong	nr_words
0	AUS	NSW	Wollondilly Shire Council	Wollondilly Shire	https://www.youtube.com/c/wollondillyshire	Road Resurfacing Video	zVr6S5XkJ28	20181127	146.120	g_NNP_2.75 'day_XX_2.75 my_PRP$_3.75 name_NN_4.53 is_VBZ_4.74 ...	62/64 Menangle St, Picton NSW 2571, Australia	(-34.1700078, 150.612913)	433
1	AUS	NSW	Wollondilly Shire Council	Wollondilly Shire	https://www.youtube.com/c/wollondillyshire	Weather update 5pm 1 March 2022 - Mayor Matt Gould	p4MjirCc1oU	20220301	181.959	hi_UH_0.64 guys_NNS_0.96 i_PRP_1.439 'm_VBP_1.439 just_RB_1.76 ...	62/64 Menangle St, Picton NSW 2571, Australia	(-34.1700078, 150.612913)	620
2	AUS	NSW	Wollondilly Shire Council	Wollondilly Shire	https://www.youtube.com/c/wollondillyshire	Transport Capital Works Video	DXlkVTcmeho	20180417	140.450	council_NNP_0.53 is_VBZ_1.53 placing_VBG_1.65 is_VBZ_2.07 2018-19_CD_2.57 ...	62/64 Menangle St, Picton NSW 2571, Australia	(-34.1700078, 150.612913)	347
3	AUS	NSW	Wollondilly Shire Council	Wollondilly Shire	https://www.youtube.com/c/wollondillyshire	Council Meeting Wrap Up February 2022	2NhuhF2fBu8	20220224	107.840	g_NNP_0.399 'day_NNP_0.399 guys_NNS_0.799 and_CC_1.12 welcome_JJ_1.199 ...	62/64 Menangle St, Picton NSW 2571, Australia	(-34.1700078, 150.612913)	341
4	AUS	NSW	Wollondilly Shire Council	Wollondilly Shire	https://www.youtube.com/c/wollondillyshire	CITY DEAL 4 March 2018	4-cv69ZcwVs	20180305	130.159	[Music]_XX_0.85 it_PRP_2.27 's_VBZ_2.27 a_DT_3.27 fantastic_JJ_3.36 ...	62/64 Menangle St, Picton NSW 2571, Australia	(-34.1700078, 150.612913)	420

CoANZSE Audio: https://coanzse.org

- Transcripts cut into 20-word chunks
- Audio segments retrieved based on word timing tags
- Searchable online CoAZNSE data, including audio and forced alignment files
- Powered by BlackLab (De Does et al. 2017), developed at the Dutch Language Institute
- "Under the hood": Apache Lucene
- Accessible via Shibboleth authentication

CoANZSE Audio: The Corpus of Australian and New Zealand Spoken English

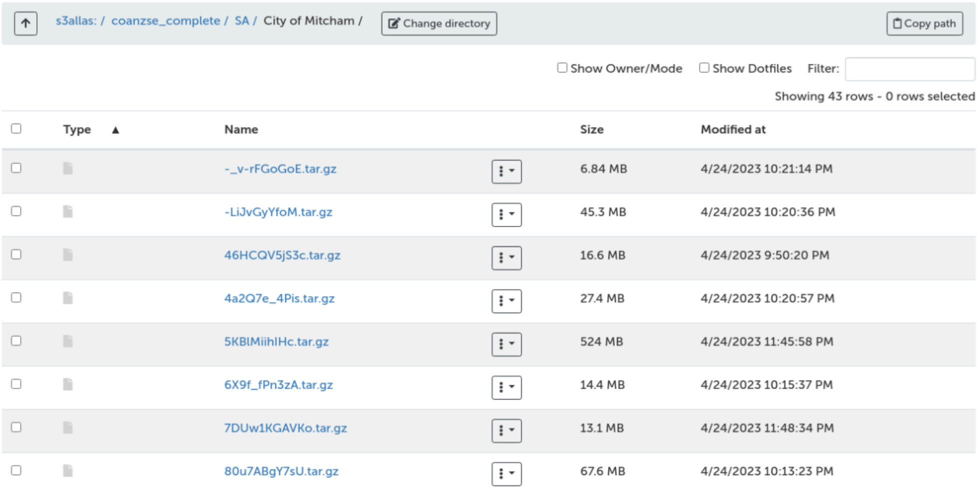
The Corpus of Australian and New Zealand Spoken English is a 195-million-word corpus of geolocated automatic speech recognition transcripts of video content from local governments in Australia and New Zealand, created for the study of lexical, grammatical, phonetic, and discourse-pragmatic phenomena of spoken language. CoANZSE Audio contains, in addition to the complete textual content of the corpus, audio files and forced alignments in Praat's TextGrid format for most transcripts.

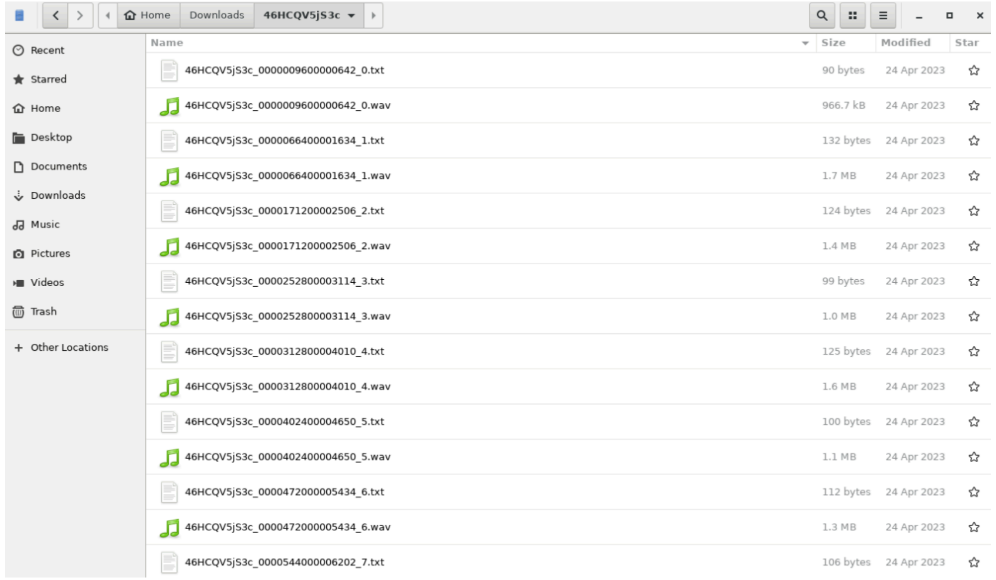
CoANZSE size by country/state/territory

Location	nr_channels	nr_videos	nr_words	video_length (h)	nr_audio_files
Australian Capital Territory	8	650	915,542	111.79	41,752
New South Wales	114	9,741	27,580,773	3,428.87	1,299,949
Northern Territory	11	289	315,300	48.72	6,628
Queensland	58	7,356	19,988,051	2,642.75	950,084
South Australia	50	3,537	13,856,275	1,716.72	643,866
Tasmania	21	1,260	5,086,867	636.99	240,453
Victoria	78	12,138	35,304,943	4,205.40	1,624,830
Western Australia	68	3,815	8,422,484	1,063.78	386,898
New Zealand	74	18,029	84,058,661	10,175.80	3,926,216
Total	**482**	**56,815**	**195,528,896**	**24,030.82**	**9,122,676**

Data workflow

Cloud storage .tar.gz files → temporary local directory → MFA → formant extraction





Pipeline: Montreal Forced Aligner (McAuliffe et al. 2017)

- For each location: all audio and transcripts (i.e. 20-word chunks) sent to MFA
- `English_MFA_Acoustic Model_v3.0.0`, trained on Common Voice English v8.0 (Ardila et al. 2020), `English_MFA.dict`
- "Adapt" functionality in MFA: Gaussian Mixture Model means adjusted on the basis of the audio for each location

```
INFO     Setting up corpus information...
INFO     Loading corpus from source files...
  0%  ─────────────────────────────────────────────── 0/100  [ 0:00:01 < -:--:-- , ? it
INFO     Found 1 speaker across 1 file, average number of utterances per speaker: 1.0
INFO     Initializing multiprocessing jobs...
WARNING  Number of jobs was specified as 3, but due to only having 1 speakers, MFA will only use 1 jobs. Use the
         --single_speaker flag if you would like to split utterances across jobs regardless of their speaker.
INFO     Normalizing text...
100%  ─────────────────────────────────────────────── 1/1  [ 0:00:00 < 0:00:00 , ? it
INFO     Generating MFCCs...
100%  ─────────────────────────────────────────────── 1/1  [ 0:00:10 < 0:00:00 , ? it
INFO     Calculating CMVN...
INFO     Generating final features...
  0%  ─────────────────────────────────────────────── 0/1  [ 0:00:01 < -:--:-- , ? it
INFO     Creating corpus split...
  0%  ─────────────────────────────────────────────── 0/1  [ 0:00:01 < -:--:-- , ? it
INFO     Compiling training graphs...
100%  ─────────────────────────────────────────────── 1/1  [ 0:00:02 < 0:00:00 , ? it
INFO     Performing first-pass alignment...
INFO     Generating alignments...
100%  ─────────────────────────────────────────────── 1/1  [ 0:00:45 < 0:00:00 , ? it
INFO     Calculating fMLLR for speaker adaptation...
100%  ─────────────────────────────────────────────── 1/1  [ 0:00:01 < 0:00:00 , ? it
INFO     Performing second-pass alignment...
INFO     Generating alignments...
  0%  ─────────────────────────────────────────────── 0/1  [ 0:00:14 < -:--:-- , ? it
```

Pipeline: Parselmouth-Praat (Jadoul et al. 2018)

- Python interface to Praat, widely used software for acoustic analysis (Boersma & Weenink 2023)
- Intergration into Python simplifies workflows and analysis
- Settings: automatic time step, five formants, maximum formant frequency of 5,500 Hz, window length 0.025 seconds, pre-emphasis > 50 Hz
- Extraction of F1 and F2 formants and bandwidths at vowel midpoint

Filtering and outliers:

- Filtering: stressed vowels in content words not followed by nasals or liquids
- Removal of stopwords with NLTK
- Word stress on the basis of values from the CMU Pronunciation Dictionary (Weide et al. 1998)
- Outliers removed with Mahalanobis distance on the basis of critical value of the 95% quantile of the χ^2 distribution

Pipeline: Spatial analysis

Tobler's first law: "everything is related to everything else, but near things are more related than distant things" (Tobler 1970)

- Moran's *I* and Getis-Ord G_i^* based on mean F1 and F2 values at each location
  - Spatial weights matrix W based on distance band $w_{ij} = 1/d_{ij}$
  - Minimum distance: all locations must have at least one neighbor
  - For each vowel and formant, only locations with at least 200 tokens considered
  - Spatial analysis conducted for AUS and NZ separately
  - Calculated with esda and pysal Python packages

Moran's *I* (Moran 1950): Takes into account attribute values at all locations in a dataset and summarizes the overall extent of spatial correlation

- 1: perfect clustering of similar values, 0: random spatial distribution of values, -1: pefectly even dispersion of values

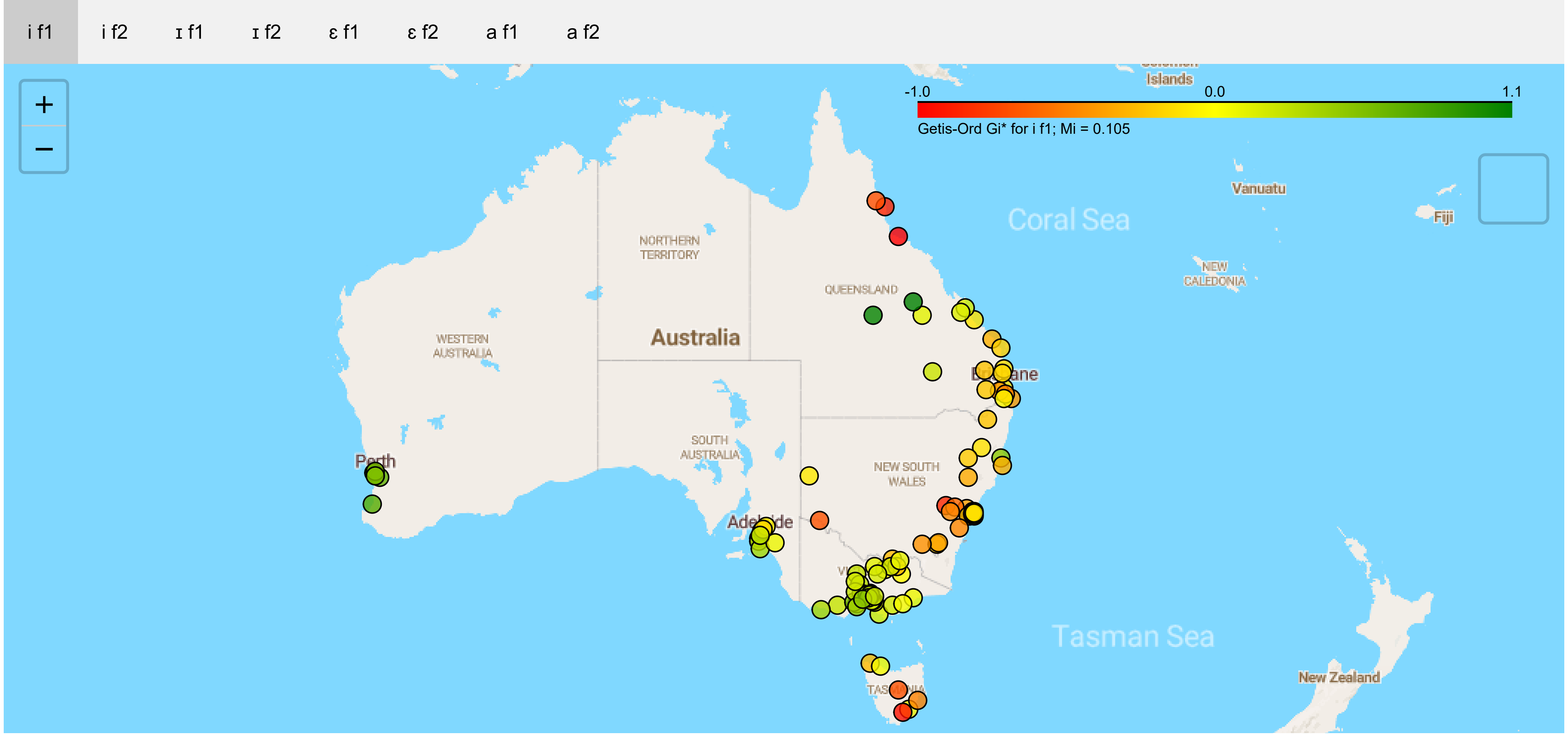
Getis-Ord G_i^* (Getis & Ord, 1992; Ord & Getis, 1995): Identifies spatial clusters by evaluating the values at each location in the dataset in comparison with neighboring locations, in relation to the global dataset

- Positive for the location and its neighbors > global mean; 0: value = global mean; negative: value < global mean

Number of vowel tokens analzyed

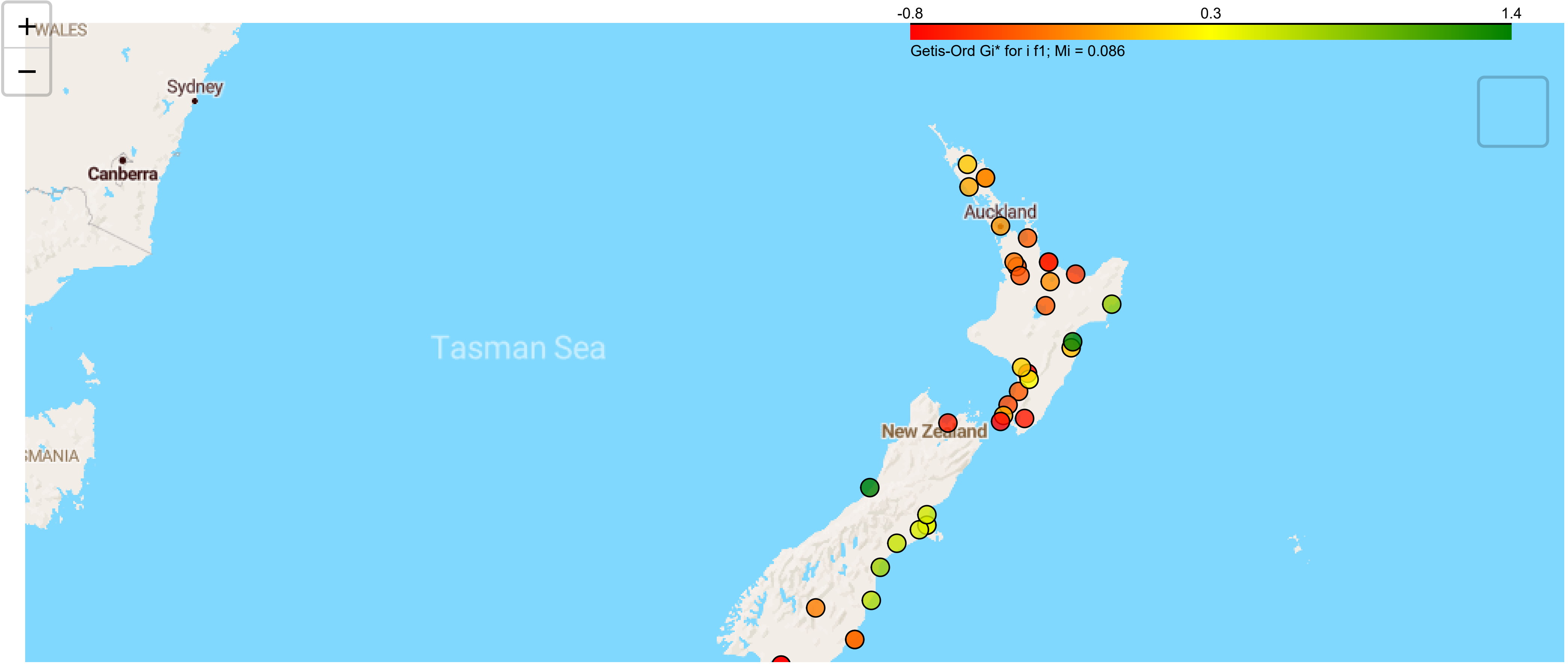
Vowel	AUS	NZ
/i/	59,950	51,189
/ɪ/	861,190	609,445
/ɛ/	1,341,498	707,105
/a/(=/æ/)	1,136,901	711,761
Total	**3,399,539**	**2,079,500**

Results: Australia



Results: New Zealand

i f1 | i f2 | ɪ f1 | ɪ f2 | ɛ f1 | ɛ f2 | a f1 | a f2



Preliminary findings

- Some evidence for overall/global spatial clustering of f1 and f2 values in Australia and New Zealand
- The magnitude of local spatial autocorrelation values is not high
- Front vowels in Australia are lowering and fronting (cf. Cox & Palethorpe 2008; Cox et al. 2024)
- This may be led by changes in the major urban centers of Sydney, Melbourne, Adelaide, and Perth
- Possibly influenced by L2 English of immigrants (Cox et al. 2024; Gonzalez et al. 2021; Travis et al. 2023)
- Possible confirmation of these findings in large-scale dataset

Caveats

- Formant extraction methods (dynamic formant tracking)
- No demographic data, but some can be semi-automatically annotated (cf. Bredin 2023; Plaquet & Bredin 2023; Ferreira 2024)
- MFA pronunciation dictionary phones differ slightly from Australian English symbols
  - Especially low and back vowels (but: Gonzalez et al. 2020; Mackenzie & Turton 2020)

Summary

- Pipeline approaches can be used to automatically process large data volumes and extract formants for millions of vowels
- Incipient regional variation attested for Australia, and (somewhat) for New Zealand
- Australian cities may be leading the changes
- Still much to be done!

References

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F., and Weber, G. (2020). Common Voice: A massively-multilingual speech corpus. *arXiv*:1912.06670 [cs.CL].

Bauer, L. and Bauer, W. (2002). Can we watch regional dialects developing in colonial English?: The case of New Zealand. *English World-Wide* 23(2), 169-193.

Boersma, P. and Weenink, D. (2024). Praat: Doing phonetics by computer.

Bredin, H. (2023). pyannote.audio 2.1 speaker diarization pipeline: Principle, benchmark, and recipe.In *INTERSPEECH 2023*, 1983-1987.

Bryant, P. (1989). The south-east lexical usage region of Australian English. *Australian Journal of Linguistics* 9, 85–134.

Horvath, B. M. and Horvath, R. J. (2001). Short A in Australian English: A geolinguistic study. In: D. Blair and P. Collins (eds.), *English in Australia*, 341-356. John Benjamins.

Coats, S. (2022). CoANZSE: The Corpus of Australian and New Zealand Spoken English: A new resource of naturalistic speech transcripts. In P. Parameswaran, J. Biggs & D. Powers (Eds.), *Proceedings of the the 20th Annual Workshop of the Australasian Language Technology Association*, 1–5.

Coats, S. (2024). CoANZSE Audio: Creation of an online corpus for linguistic and phonetic snalysis of Australian and New Zealand Englishes. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 3407-3412.

Cox, F. and Palethorpe, S. (2019). Vowel variation in a standard context across four major Australian cities. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* pp. 577-581.

Cox, F. and Palethorpe, S. (2008). Reversal of short front vowel raising in Australian English. *Interspeech 2008* pp. 342-345.

Cox, F., Penney J., and Palethorpe, S. (2024). Australian English monophthong change across 50 Years: Static versus dynamic measures. Languages, 9(3).

Ferreira, A. I. S. (2024). wav2vec2-large-xlsr-683 53-gender-recognition-librispeech.

Getis, A. and Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24, 189-206.

Gonzalez, S., Grama, J., and Travis, C. E. (2020). Comparing the performance of forced aligners used in sociophonetic research. *Linguistics Vanguard* 6, 20190058.

Grama, J., Travis, C. E., and Gonzalez, S. (2021). Ethnic variation in real time: Change in Australian English diphthongs. In H. Van de Velde, N. H. Hilton, and R. Knooihuizen (eds.), *Language Variation--European Perspectives VIII*, 291-314. John Benjamins

Horvath, B. M. and Horvath, R. J. (2001). Short A in Australian English: a geolinguistic study. In D. Blair and P. Collins (eds.), *English in Australia*, 341-356. John Benjamins.

Jadoul, Y., Thompson, B., and de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics* 71, 1-15.

Kennedy, M. (2006). *Variation in the pronunciation of English by New Zealand school children*. Master's thesis, Victoria University of Wellington.

MacKenzie, L. and Turton, D. (2020). Assessing the accuracy of existing forced alignment software on varieties of British English. *Linguistics Vanguard* 6.

Marsden, S. (2013). *Phonological variation and the construction of regional identities in New Zealand English*. Doctoral thesis, Victoria University of Wellington.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association*, 498-502.

Mitchell, A. G. and Delbridge, A. (1965). *The speech of Australian adolescents: A survey*. Angus and Robertson.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika* 37, 17-23.

Ord, J. K. and Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* 27, 286-306.

Plaquet, A. and Bredin, H. (2023). Powerset multi-class cross entropy loss for neural speaker diarization. In *INTERSPEECH 2023*, 3222-3226.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46, 234-240.

Travis, C. E., Grama, J., Gonzalez, S., Purser, B., and Johnstone, C. (2023). Sydney Speaks Corpus.

Weide, R. and others (1998). The Carnegie Mellon Pronouncing Dictionary.