

Phonetic Segmentation of the UCLA Phonetics Lab Archive

Eleanor Chodroff¹, Blaž Pažon²,
Annie Baker¹ & Steven Moran^{3,4}

¹University of Zurich, ²University of Ljubljana, ³University of Neuchâtel,
⁴University of Miami

Access the VoxAngeles Corpus

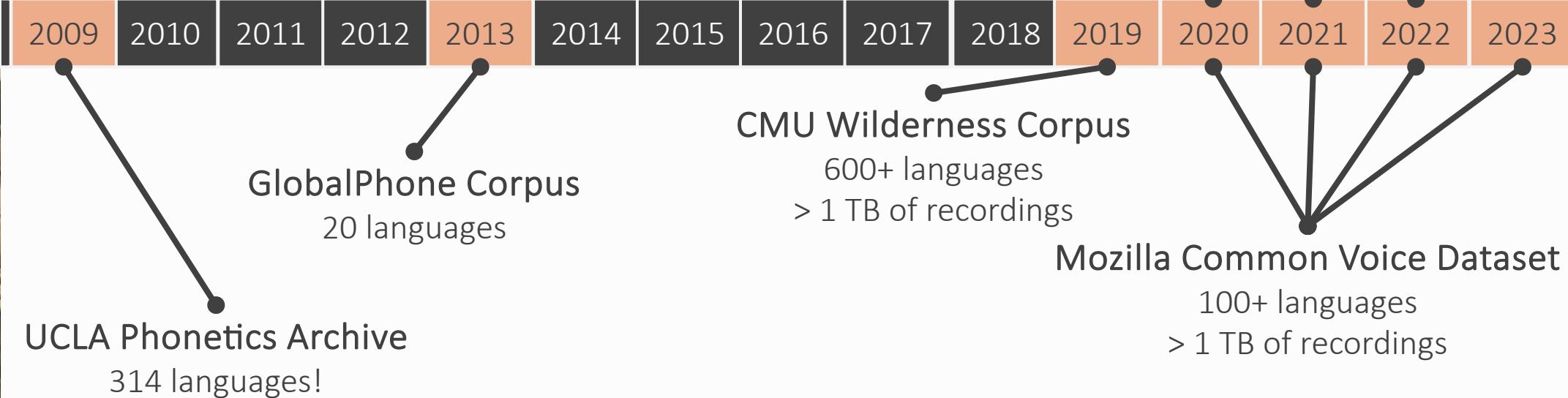


github.com/pacscilab/voxangeles





Multilingual speech corpora

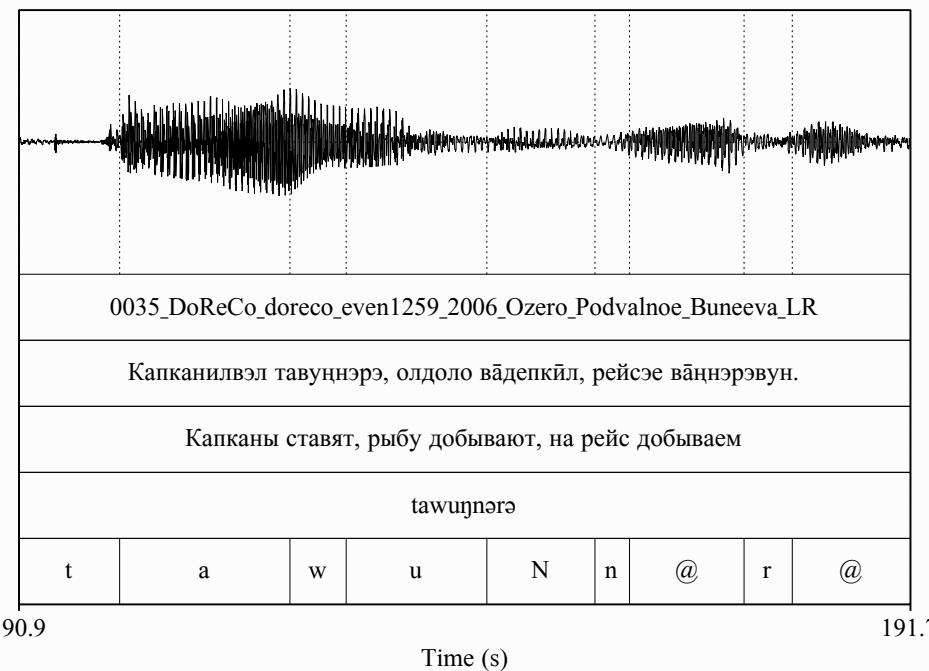


Applications for speech technology
Advances in speech science and phonetic theory



But what makes a corpus usable?

- Time-aligned transcriptions: utterance-, word-, and phone-levels
- Clean and well-annotated data: correct transcript and phonetic transcriptions





Goals

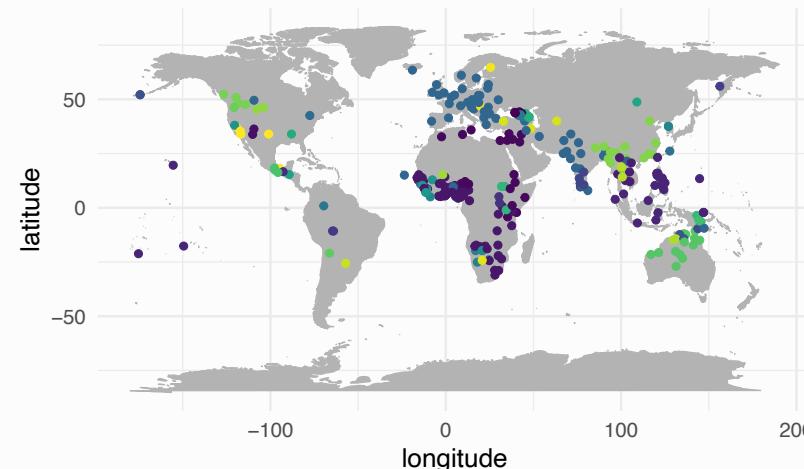
- Further enhance the usability of the UCLA Phonetics Lab Archive
- Word-level alignments → phone-level alignments
- Standardize transcription
- Extract phonetic measurements
- Demonstrate usability for phonetic typology → intrinsic f0 case study

VoxAngeles Corpus



UCLA Phonetics Lab Archive

- The original massively multilingual speech corpus
- 314 languages from 51 language families
- Spoken audio recordings & phonetic transcriptions (not time-aligned)



UCLA Phonetics Lab Archive

1960s – late 1990s: data collection in fieldwork sessions

Most commonly word list data

continued on
Moscow Tape 2. Side 1 track 3-2
 (013) [Shota Arustavi]
 JCC · Afrasiani 33uk - the numbers one to twenty.

1-5	a'k ^z ə y:bæ	y:bæ	x:p ^z ə pʃ:bæ	pʃ:bæ
6-10	f:bæ	bʒ:bæ	tä:b ^z	x:bæ
11-15	'ye:z ^z	y ^z ʒy ^z	yaxə	yib ^z
16-20	yaf	yib ^z ^z	ya:	zey ^z y ^z ya

1. ə/ʃ/ʃ/ʃ/
 2. b^z/ʃ/ʃ/ʃ/
 3. ʃ/ʃ/
 4. ʃ/ʃ/
 5. ʃ/ʃ/
 6. ʃ/ʃ/
 7. ʃ/ʃ/
 8. ʃ/ʃ/
 9. ʃ/ʃ/
 10. ʃ/ʃ/
 11. ʃ/ʃ/
 12. ʃ/ʃ/
 13. ʃ/ʃ/
 14. ʃ/ʃ/
 15. ʃ/ʃ/
 16. ʃ/ʃ/
 17. ʃ/ʃ/
 18. ʃ/ʃ/
 19. ʃ/ʃ/
 20. ʃ/ʃ/
 21. ʃ/ʃ/
 22. ʃ/ʃ/
 23. ʃ/ʃ/
 24. ʃ/ʃ/
 25. ʃ/ʃ/
 26. ʃ/ʃ/
 27. ʃ/ʃ/
 28. ʃ/ʃ/
 29. ʃ/ʃ/
 30. ʃ/ʃ/
 31. ʃ/ʃ/
 32. ʃ/ʃ/
 33. ʃ/ʃ/
 34. ʃ/ʃ/
 35. ʃ/ʃ/
 36. ʃ/ʃ/
 37. ʃ/ʃ/
 38. ʃ/ʃ/
 39. ʃ/ʃ/
 40. ʃ/ʃ/
 41. ʃ/ʃ/
 42. ʃ/ʃ/
 43. ʃ/ʃ/
 44. ʃ/ʃ/
 45. mentón mawápə
 46. barba hák=yúk
 47. cuello bít=na?
 48. estómago wída?
 49. espalda tā?
 50. hombro dad?
 51. brazo mihna?
 52. antebrazo mihná?
- what is g? ↓
 what is the difference between [x̥] and [x]? →
 what is the difference between [ʃ] and [χ]? →
1. ə/ʃ/ʃ/ʃ/
 2. b^z/ʃ/ʃ/ʃ/
 3. ʃ/ʃ/
 4. ʃ/ʃ/
 5. ʃ/ʃ/
 6. ʃ/ʃ/
 7. ʃ/ʃ/
 8. ʃ/ʃ/
 9. ʃ/ʃ/
 10. ʃ/ʃ/
 11. ʃ/ʃ/
 12. ʃ/ʃ/
 13. ʃ/ʃ/
 14. ʃ/ʃ/
 15. ʃ/ʃ/
 16. ʃ/ʃ/
 17. ʃ/ʃ/
 18. ʃ/ʃ/
 19. ʃ/ʃ/
 20. ʃ/ʃ/
 21. ʃ/ʃ/
 22. ʃ/ʃ/
 23. ʃ/ʃ/
 24. ʃ/ʃ/
 25. ʃ/ʃ/
 26. ʃ/ʃ/
 27. ʃ/ʃ/
 28. ʃ/ʃ/
 29. ʃ/ʃ/
 30. ʃ/ʃ/
 31. ʃ/ʃ/
 32. ʃ/ʃ/
 33. ʃ/ʃ/
 34. ʃ/ʃ/
 35. ʃ/ʃ/
 36. ʃ/ʃ/
 37. ʃ/ʃ/
 38. ʃ/ʃ/
 39. ʃ/ʃ/
 40. ʃ/ʃ/
 41. ʃ/ʃ/
 42. ʃ/ʃ/
 43. ʃ/ʃ/
 44. ʃ/ʃ/
 45. mentón mawápə
 46. barba hák=yúk
 47. cuello bít=na?
 48. estómago wída?
 49. espalda tā?
 50. hombro dad?
 51. brazo mihna?
 52. antebrazo mihná?

UCLA Phonetics Lab Archive

2004 – 2009: Digitization of phonetic transcriptions, orthographic forms and translations

Abkhaz

Note: Some of the Word Lists below contain special characters in the Mkhedruli script. To view these characters, please download and install the TITUS Cyberbit Basic font, available free for Mac, Linux, and Windows at the [TITUS Project homepage](#).

[Back to Language Index](#)

Consult the Word Lists in order to locate individual words of interest. You can search within your web-browser for specific sounds. (See word lists for more details.)

Left-click to access files online; right-click (Macintosh Control + click) to download. Not all information is available for all recordings, and so some cells might be blank.

Click "Details" for more information about a specific recording.

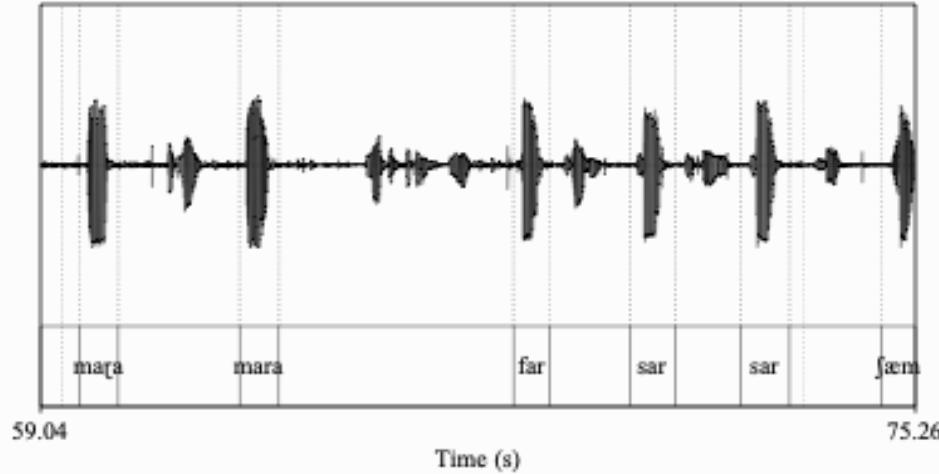
For more detailed instructions, click [here](#).

	Word List	Word List Entries	Additional Info	Audio Filename	WAV	MP3	Scanned Word List (JPG)	JPG 2	Scanned Word List (TIF)	TIF 2	Recording Details
1	abk_word-list_1970_01.html	1 - 20	Speaker origin unknown; dialect unspecified	abk_word-list_1970_01	WAV	MP3	JPG		TIF		Details
2	abk_word-list_1970_02.html	1 - 20	Speaker origin unknown; Bzyb dialect	abk_word-list_1970_02	WAV	MP3	JPG		TIF		Details
3	abk_word-list_1970_03.html	1 - 44	Speaker origin unknown; Bzyb dialect	abk_word-list_1970_03	WAV	MP3	JPG	JPG 2	TIF	TIF 2	Details
4	abk_story_1970_01.html	1 - 9	Speaker origin unknown; Bzyb dialect	abk_story_1970_01	WAV	MP3					Details
5	abk_word-list_1977_01.html	1 - 100	Speaker from Lykhny, Gudauta District, Abkhazia; Bzyb dialect	abk_word-list_1977_01	WAV	MP3	JPG	JPG 2	TIF	TIF 2	Details

Entry	Abkhaz transcription	English
1	a'kə	one
2	y:'bæ	two
3	χ:pʰæ	three
4	pç:'bæ	four
5	χʷ:bæ	five
6	fɪbæ	six
7	bz:'bæ	seven
8	?ä:'bʒ	eight
9	z̥w:’bæ	nine
10	z̥wʒ'bæ	ten
11	lɔ:w'e'zʒ	eleven
12	z̥wøqʒ	twelve

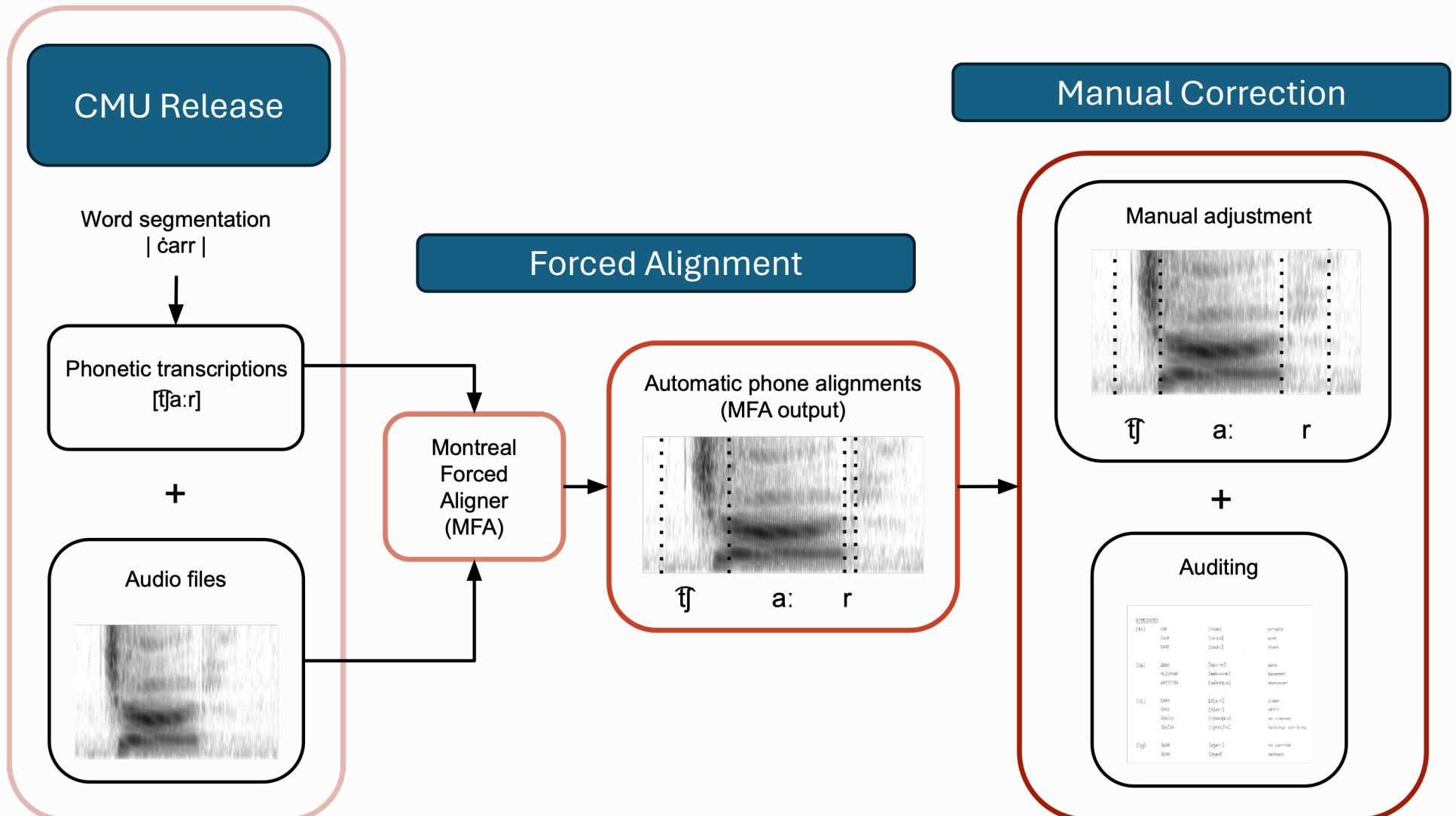
CMU Release

- 95 languages, 21 language families
- Word-level audio files and transcripts using an automated approach
- Audio segmentations validated by a human listener to ensure correct alignment
- Greatly enhanced the usability of the corpus!



gaa-001-008	e	d	z ^w	a
gaa-001-009	e	d	z	a
gaa-001-010	e	k	a	
gaa-001-011	e	q	ɔ	i
gaa-001-012	e	m	a	
gaa-001-013	e	ŋ	m	o
gaa-001-014	e	n	a	
gaa-001-015	e	ŋ	a	
gaa-001-016	e	ŋ	a	
gaa-001-018	e	ɥ	e	r e l a
gaa-001-019	e	w	e	

Methods Overview





Forced alignment

- Montreal Forced Aligner with the english 2.0.0a “Global English” acoustic model
 - Trained on 3700+ hours of >6 English varieties
 - Competitive with small language-specific acoustic models in forced alignment performance (Chodroff et al. 2024)
 - Mapped crosslinguistic phone set to Global English phone set using the Interlingual MFA toolkit (Dolatian 2023)
 - Mapped back to original phone set after alignment



Manual correction

Manual auditing of transcriptions
Manual alignment of boundaries

Principles:

- **Auditing:** Remain faithful to the original linguist's transcription
- **Alignment:** Represent the entire speech signal with the provided transcription
- **Alignment:** Assign a section of the speech signal to each element in the transcription

~ 30% of original phonetic transcriptions updated across 85 languages
~71% of the force-aligned boundaries were within 20 ms of the gold boundary



Manual auditing

- Consulted the original field notes (images) in most cases to verify the transcription
- Obsolete and non-standard symbols were updated
- Missing IPA symbols were re-added where needed
- Example: Ganda (lug) and Lendu (led): <dd> represented phonetically distinct segments:
 - In Ganda: geminated, or long, consonant - updated to <d:>
 - In Lendu: a non-pulmonic implosive consonant - updated to <d>

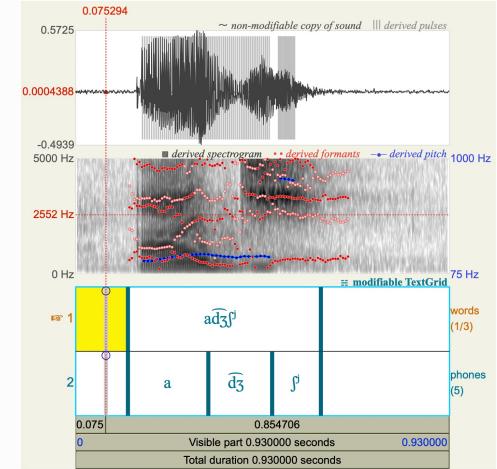




VoxAngeles:

github.com/pacscilab/voxangeles

- 95 languages, 21 language families, 5445 recordings (median 49 per language)
- Additional 11 languages extracted and force-aligned but not yet audited
- Phone-aligned TextGrids with phonetic transcriptions and corresponding text files
- Text files with initial and updates transcriptions
- Pre-extracted acoustic phonetic measurements:
 - Duration of all phone segments
 - Vowel f0 at each quartile and decile
 - Vowel F1 - F3 at each quartile and decile



lang	file	raw	intermediate1	intermediate2	updated
abk	abk-002-000	a-dʒʃɪ	a ə dʒɪʃɪ	a ə dʒɪʃɪ	a ə dʒɪʃɪ
abk	abk-002-001	'a-dʒməz	a ə dʒɪ m ə z	a ə dʒɪ m ə z	a ə dʒɪ m ə z
abk	abk-002-006	adʒəmʃə	a ə dʒɪ m əʃə	a ə dʒɪ m əʃə	a ə dʒɪ m əʃə
abk	abk-002-009	atʃərā	a tʃɪ r ə	a tʃɪ r ə	a tʃɪ r ə
abk	abk-002-010	átfəpʰərʌ	a tʃɪ pʰ ə r ʌ	a tʃɪ pʰ ər ʌ	a tʃɪ pʰ ər ʌ
abk	abk-002-011	áttʃʃərə	a tʃɪʃʃ ə r ə	a tʃɪʃʃ ə r ə	a tʃɪʃʃ ə r ə
abk	abk-002-023	ak'áʃərə	a k' a əʃ ə r ə	ak'áʃərə	ak'áʃərə
abk	abk-002-024	ăbʒɪʃ	ă b əʒɪʃ	ăbʒɪʃ	ăbʒɪʃ
abk	abk-002-026	aʃæ	a ʃ æ	a ʃ æ	a ʃ æ
abk	abk-002-027	ájəʃɛ	a j əʃ ɛ	a j əʃ ɛ	a j əʃ ɛ
abk	abk-002-028	'aʃæ	a ʃ æ	a ʃ æ	a ʃ æ



Real-world applications

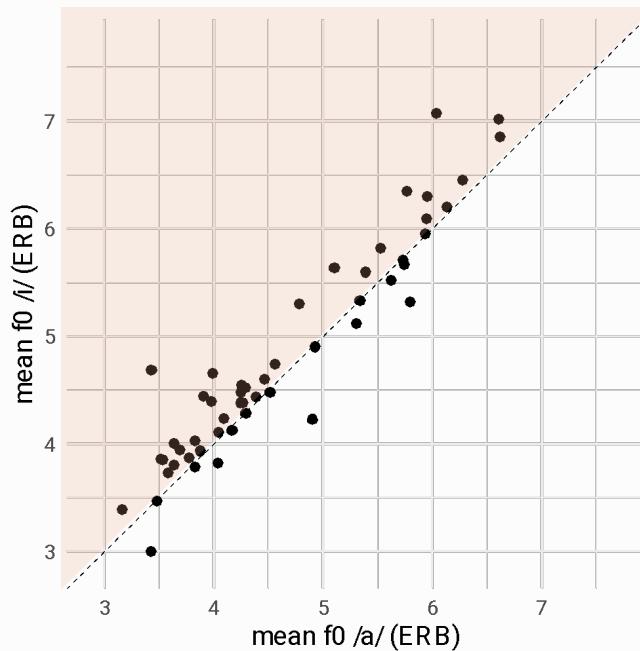
- Pedagogical purposes
- Research in multilingual phone recognition and alignment
- Research in phonetic typology
 - Ex: Intrinsic vowel f0, intrinsic vowel duration, or consonant f0



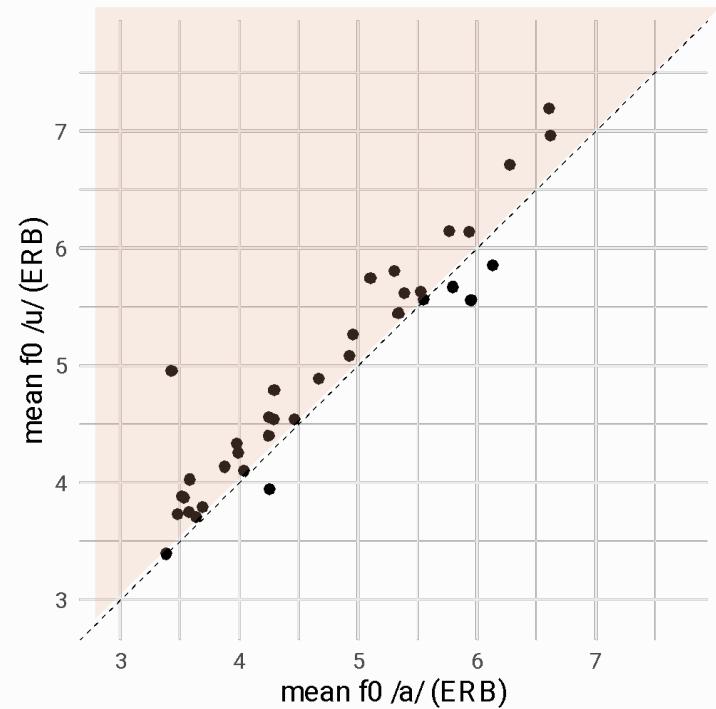
Case study: Intrinsic f0

- Do high vowels (e.g. /i/ /u/) have a higher f0 than low vowels (e.g. /a/)?
- Previous confirmation of the effect:
 - Whalen and Levitt (1995) with 31 languages
 - Ting et al. (2023) with 16 languages but more data per language
- Vowels were collapsed into three broad categories:
 - High front vowels: /i/ - includes [i ɪ ɿ ɿ]
 - High back vowels: /u/ - includes [ʊ ɻ ɯ ɯ]
 - Low vowels: /a/ - includes [a ɑ æ ɛ]
- Extracted midpoint f0, converted to ERB (Glasberg and Moore 1990)

Case study: Results



/i/–/a/
52 languages
71% conformity (31/52)
 $r = 0.95$



/u/–/a/
37 languages
86% conformity (32/37)
 $r = 0.95$



Case study: Results

$f_0 \text{ midpoint (ERB) } \sim$
vowel height * preceding voice * following voice +
vowel duration +
(1 + vowel height | language)

- Both /i/–/a/ and /u/–/a/ models:
 - **High vowel f0 > low vowel f0 (IFO)**
 - $f_0 \text{ following vcl} > f_0 \text{ following vcd } (\sim CFO)$
 - $f_0 \text{ at the end of a word} < f_0 \text{ before segment}$
 - $f_0 \text{ increases with vowel duration}$
- /i/–/a/:
 - $f_0 \text{ difference by height smaller at the end of a word}$
- /u/–/a/:
 - $f_0 \text{ difference by height larger after vcl than vcd}$
 - $f_0 \text{ difference by height larger before vcl than vcd}$



Future directions

- Continue data extraction from the UCLA Phonetics Lab Archive to expand the VoxAngeles corpus, both in terms of number of languages and speakers per language
- Segmentation of data from unprocessed languages - this will further diversify the available data for analysis
- Currently hosted on Github (github.com/pacscilab/voxangeles), but perhaps also host on original UCLA website



Acknowledgments

Thank you to:
Ryan Cotterell
Claire Moore-Cantwell
Pat Keating
Ian Maddieson



SNF Grant PROOP1_208460 to EC
SNF Grant PCEFP1_186841 to SM



References

- Chodroff, E., Ahn, E. P., Dolatian, H.. (2024). Comparing cross-language and language-specific acoustic models for low-resource phonetic forced alignment. *Language Documentation and Conservation*.
- Dolatian, H. (2023). Interlingual MFA. GitHub. <https://github.com/jhdeov/interlingual-MFA>
- Glasberg, B. R. and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103–138.
- Ladefoged, P., Blankenship, B., Schuh, R. G., Jones, P., Gfroerer, N., Griffiths, E., Harrington, L., Hipp, C., Jones, P., Kaneko, M., Moore-Cantwell, C., Oh, G., Pfister, K., Vaughan, K., Videc, R., Weismuller, S., Weiss, S., White, J., Conlon, S., Lee, WS. J., and Toribio, R. (2009). The UCLA Phonetics Lab Archive. Los Angeles, CA: UCLA Department of Linguistics. <http://archive.phonetics.ucla.edu>
- Li, X., Mortensen, D. R., Metze, F., & Black, A. W. (2021). Multilingual phonetic dataset for low resource speech recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6958-6962). IEEE. Retrieved July 2023.
- McAuliffe, M. And Sonderegger, M. (2022). English MFA acoustic model v2.0.0. <https://mfa-models.readthedocs.io/acoustic/English/> English MFA acoustic model v2_0_0.html
- Ting, C., Clayards, M., Sonderegger, M., and McAuliffe, M. (2023, March 24). The cross-linguistic distribution of vowel and consonant intrinsic F0 effects.
- Whalen, D. H. and Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23(3), 349–366.