Autophon user guide
Spanish – America

Engine: faseAlign
Model: Release 1.1.14

1 Introducing Autophon and Forced Alignment

Autophon is a free, user-friendly tool for phoneticians that performs *forced alignment* (FA) — the automated process of converting speech recordings and their transcriptions into phonetically time-stamped annotations.

Autophon leverages widely used alignment engines developed by the phonetics community, including:

- *FAVE*[1]
- *faseAlign*[2]
- *Montreal Forced Aligner, version 1.0*[3]
- *Montreal Forced Aligner, version 2.0*[3]

The tool produces time-aligned phonetic annotations compatible with Praat[4], based on two user inputs: (1) a speech recording and (2) its orthographic transcript.

This user guide is specifically for ***Spanish***, using the ***faseAlign*** engine with the ***Release 1.1.14*** model. Autophon may support additional engine-model combinations for this language; therefore, ensure you are using the best option for your needs.

While many forced aligners exist, they often require command-line usage and are tied to outdated or incompatible operating systems. **Autophon offers a platform-independent, intuitive alternative for phoneticians worldwide**.

2 Using the app

2.1 Aligning files without registering

To align smaller files, go to the main page and click *Add files* at the bottom. A box titled *Transcription Mode: change transcription mode* will appear. Click the heading to choose one of four *Transcription Modes* (see below), then select your files.

2.2 Registering and logging in

To align larger files or access the full suite, click *Sign up* to create a free account. This helps us monitor usage for funders and guard against bots. After signing up, check your email for a verification link. If it doesn't arrive, check your spambox and wait 15 minutes before contacting tech support.

2.3 Cost

Autophon is free of charge.

2.4 Aligning files in a registered account

Once registered and verified, go to the *Aligner* tab and click *Add files*. A box titled *Transcription Mode: change transcription mode* will appear. Click the heading to choose one of four *Transcription Modes*, then select your files.

2.5 Transcription modes

Autophon supports four *Transcription Modes*, named for the fields they're commonly used in: *Experimental Linguistics A*, *Experimental Linguistics B*, *Computational Linguistics*, and *Variationist Linguistics*. Each mode can be selected via the corresponding box in Figure 1, which illustrates expected file structures and links to instructional videos.

Video instructions for each transcription mode can be viewed. In addition, sample templates for each mode are available for **download here**.

[1] FAVE was built by Rosenfelder, Fruehwald, Brickhouse, Evanini, Seyfarth, Gorman, Prichard, and Yuan (2022). It relies on the Hidden Markov Toolkit (Young, Woodland, and Byrne 1993).

[2] faseAlign was built by Wilbanks (2022). Like FAVE, it relies on the Hidden Markov Toolkit (Young, Woodland, and Byrne 1993).

[3] The Montreal Forced Aligner was developed by McAuliffe, Socolof, Mihuc, Wagner, and Sonderegger (2017). It uses the Kaldi toolkit (Povey, Ghoshal, Boulianne, Burget, Glembek, Goel, Hannemann, Motlicek, Qian, Schwarz, et al. 2011).

[4] Praat is a speech analysis tool developed by Boersma and Weenink (2017).

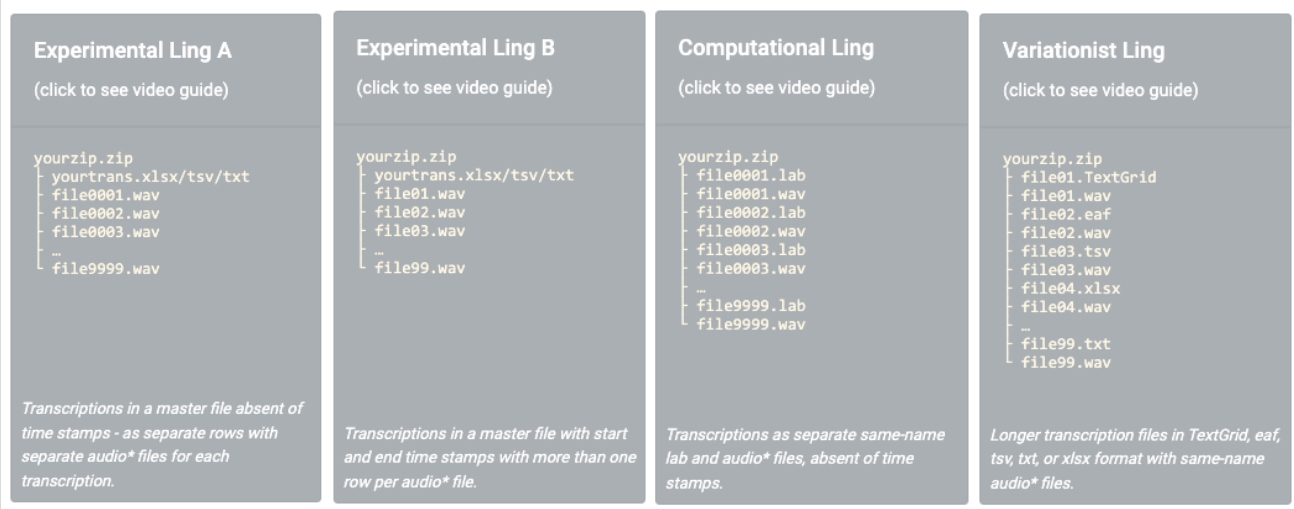


Figure 1: The Transcription Mode selection menu for Autophon.

Figure 2: Autophon outputs finished TextGrids using the same subfolder structure as the uploaded files.

Experimental linguistics A: Upload a two-column spreadsheet (Excel `xlsx`, or tab-delimited `txt`/`tsv`) with audio filenames in column 1 and transcriptions in column 2. No time stamps allowed. This format suits short clips and resembles CommonVoice[5]. Use zip or individual file upload.

Experimental linguistics B: Same structure as A, but with four columns: audio file name, start time, end time, and transcription. Designed for longer recordings requiring segmentation. Time stamps must be in real-number seconds (e.g., `1.23` or `1,23`); no colons or hour-minute markers are permitted (e.g., you may not use something like `00:00:01.23`).

Computational Linguistics: Upload matching audio and `lab` files (containing only the corresponding transcription). Files may be zipped with nested folders—Autophon preserves the hierarchy (Figure 2). No time stamps permitted.

Variationist Linguistics:[6] Upload paired transcription and audio files (individually or zipped). Transcriptions may be in Praat `TextGrid`, ELAN `eaf`, or tabular format (`xlsx`, `txt`, `tsv`). Use either three or four columns:

- Four-column: speaker, start time, end time, transcription
- Three-column: start time, end time, transcription

Time stamps must be real-number seconds (comma or period decimal separators); formats with colons (e.g., `00:00:01.23`) are not supported.

2.6 File formats and codecs

If you encounter errors during upload, it's often due to unsupported file formats or *codecs*. The simplest fix is to re-save your files in a common format using tools like Praat or ELAN.

Transcription file formats: Autophon accepts transcription files in most standard encodings, including `UTF-8` and `UTF-16 (Windows CRLF)`. If you encounter issues, try re-saving the file or email a sample to tech support.

Audio file formats: Autophon supports a wide range of audio formats, including: `WAV`, `FLAC`, `MP3`, and more. Stereo files are not currently accepted. Therefore, convert all audio to mono first.

2.7 Transcription preparation

Regardless of the transcription mode, each entry should contain between one and 20 words. Boundary demarcations must include at least 0.01 seconds of silence before and after the speech. Figure 3 shows a five-word phrase with a 0.03-second pre-boundary and a 0.25-second post-boundary. This sort of variability is expected and handled well by Autophon.[7]

2.8 Select a language

After uploading files into the aligner, Autophon will suggest a language and language model. You may override this suggestion using the dropdown menu.

2.9 Task list

The task list displays all uploads along with file name, upload date, language, tier count, file size, word count, and an inventory of missing words. You can delete the task and start over, add words to your *custom pronunciations* box (described below), or proceed by clicking *Align*.

2.10 Missing words

To understand this feature, it helps to know that forced alignment maps phonemic pronunciations – defined in language-specific dictionaries – onto the speech stream using statistical models. These dictionaries contain a finite set of words. The *missing words* feature lists items not found in Autophon's dictionary and provides suggested pronunciations. Autophon will use these suggestions by default, but you can reject them and enter your own. The next section explains how.

2.11 Your custom pronunciations

If you disagree with either (a) Autophon's pronunciation suggestions for missing words or (b) the default dictionary entries, you can override them here. Enter your own phonemic transcriptions in this box, which will take precedence over both.

Pronunciations must be entered using the alphanumeric string specific to the language model at hand – in this case, the ***faseAlign phone set***. Section 4 provides a key that maps the faseAlign phone set to its IPA[8] equivalents.

[5] https://commonvoice.mozilla.org

[6] This field originally drove the development of forced alignment in the early 2000s.

[7] If your transcriptions are segmented with *exact* start and end times, performance may degrade and boundary shifts may occur. If you're working with such data, contact tech support—we are interested in designing a fifth transcription mode for these cases.

[8] International Phonetic Alphabet

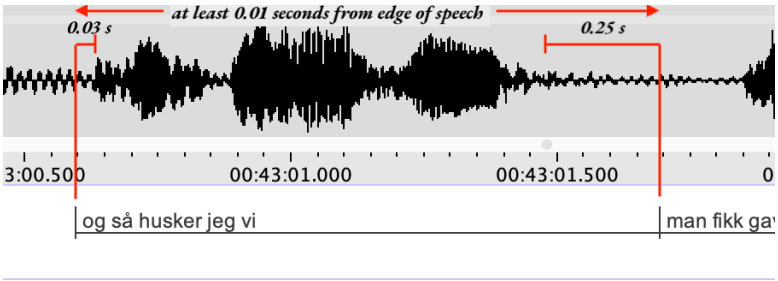


Figure 3: A sample transcription with at least 0.01 seconds of buffer on either end of the speech stream.

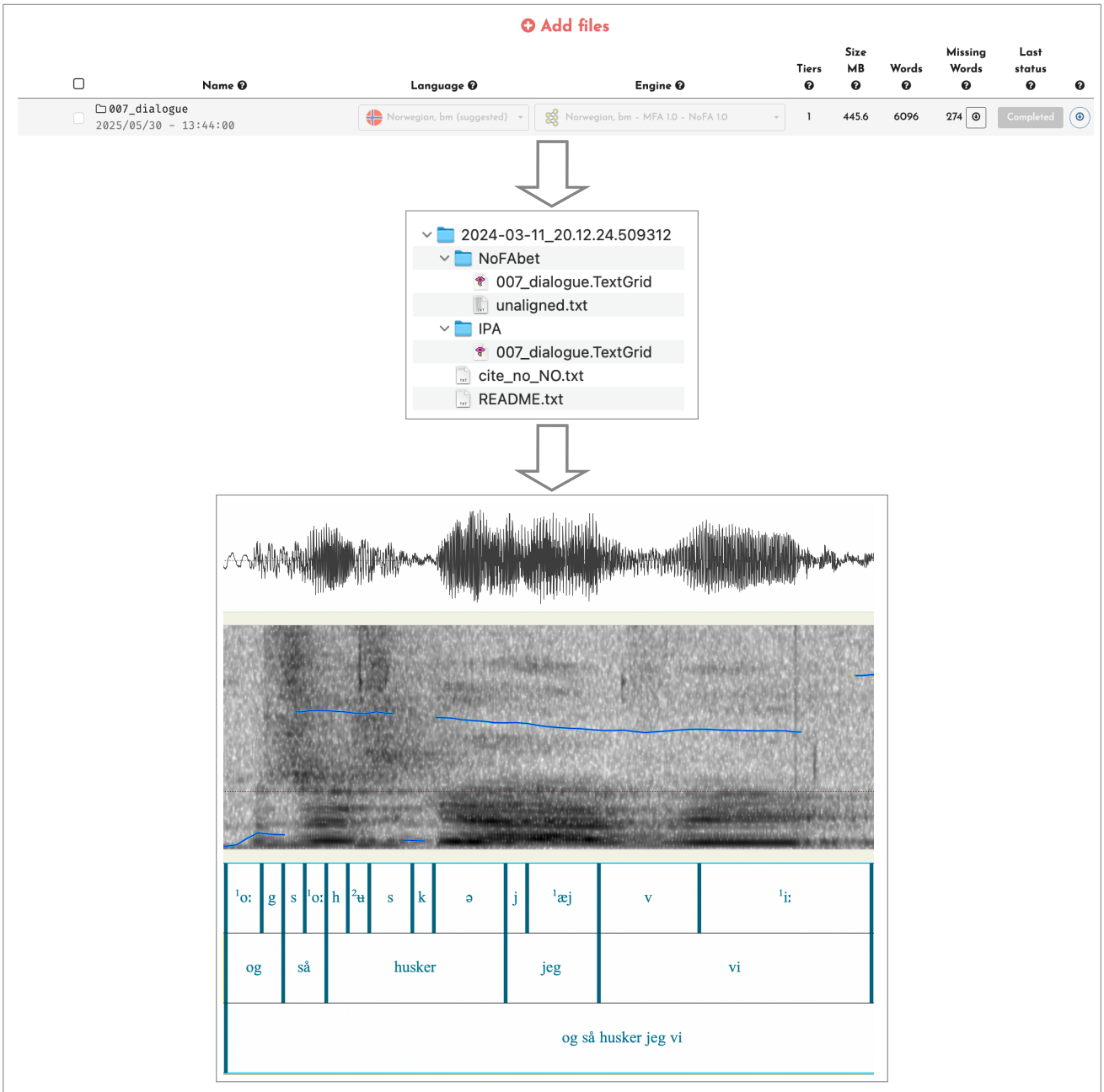


Figure 4: The alignment process, including task list, folder structure, and Praat TextGrid.

You can enter pronunciations directly in the dictionary box or upload them from a `txt` file. The maximum input length is 50 000 characters.

Entries must follow the format:

- `word[space]phoneme[space]phoneme` OR `word[tab]phoneme[space]phoneme`

Each phoneme must be separated by a space, and the lookup may not include two or more words — Autophon will interpret the second word as a phone and produce an error. You may submit multiple pronunciations for the same word by repeating the word on separate lines with different phoneme strings. Autophon will evaluate the best match for each speech instance. Refer to Figure 5 and the examples below.

Correct:	`congratule`	`k N g R a t y l`
	`congratule`	`k N R a t y l`
	`con_gratule`	`k N g R a t y l`
Incorrect:	`congratule`	`k N gR a t y l` (phones missing a space between them)
	`con gratule`	`k N g R a t y l` (two look-ups on a single line)

Figure 5: Interface with dictionary entry (left) and phoneme string input (right).

2.12 Aligning files To begin alignment, click *Align* to the far right of the upload list. Alignment typically takes a few minutes, depending on server load.

2.13 Downloading the annotations When alignment is complete, you can download the annotations as Praat TextGrids by clicking the downward arrow beside the task list. See Figure 4 for an illustration.

3 How to cite

Any dissemination or publication that makes use of this Autophon package for ***Spanish – America***, which uses ***Release 1.1.14*** within ***faseAlign*** for its engine, should cite the relevant references listed below. Proper citation is essential: not only to acknowledge the "daisy chain" of technical and academic work underpinning Autophon, but also to reinforce the incentives for sharing tools with the broader community.

While space constraints may tempt you to remove references to software, we strongly encourage prioritizing these citations. If trimming is necessary, consider reducing peripheral citations in the literature review instead.

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [softw.], ver.6.0.36. www.praat.org

Wilbanks, E. (2022). faseAlign (Version 1.1.14). https://github.com/EricWilbanks/faseAlign

Young, N. J., & Anikwe, K. H. (2025). Autophon – Automatic phonetic annotation and online forced aligner. www.autophon.org

4 Phoneme key

Autophon will output two versions of the same TextGrid for every file you align: (1) a TextGrid in the faseAlign phone set specific to Release 1.1.14 for faseAlign and (2) a TextGrid in the International Phonetic Alphabet (IPA). The phoneme key is located in Table 1.

We encourage you to inform us of errors and provide suggestions for changes.

faseAlign phone set	*IPA*	*example*	*faseAlign phone set*	*IPA*	*example*	*faseAlign phone set*	*IPA*	*example*
Vowels			*Consonants*			x	x	*Jorge*
i	i	*mismo*	p	p	*pan*	Y	j	*yo*
u	u	*tú*	b	b	*boca*	CH	tʃ	*chico*
o	o	*todo*	t	t	*taza*	m	m	*mamá*
e	e	*gente*	d	d	*día*	n	n	*noche*
a	a	*casa*	k	k	*calle*	NY	ɲ	*niño*
			g	g	*gato*	l	l	*lunes*
			f	f	*fruta*	r	ɾ	*pero*
			s	s	*sol*	R	r	*perro*

Table 1: faseAlign phone set, IPA, and lexical examples.

5 Acoustic model and pronunciation dictionary

This specific Autophon package for ***Spanish – America*** uses Release 1.1.14 within faseAlign, which "was trained using the HTK on data from 20 spontaneous interviews from the Corpus del Español de Raleigh-Durham (CERD) created and maintained by Drs. Jim Michnowicz and Rebecca Ronquest."[9].

6 Performance metrics

Metrics for this specific model can be accessed in Wilbanks (2015) (attached as an Appendix here).

7 Data security and GDPR compliance

Files uploaded to Autophon are encrypted and transmitted to a secure server hosted by *Digital Ocean* within the European Union (Frankfurt and Amsterdam). Transcriptions and audio files are automatically deleted immediately after alignment. This approach enhances privacy while also reducing storage costs. By contrast, finished TextGrids remain available in your account until you choose to delete them. Once deleted, they are permanently removed from our servers.

If you upload files but do not initiate alignment by clicking *Align*, the files will be automatically purged at 3:00 AM GMT[10].

Autophon adheres to the principles of the European Union's General Data Protection Regulation (GDPR). We collect only four pieces of user information: name, title, affiliation, and email address. Once a file is aligned, the corresponding audio is permanently deleted. Deleting a file from your task list also permanently removes the transcription and filename metadata. You may delete your account at any time, which will erase all associated personal data. However, we **do** retain anonymized alignment metadata — such as a randomly assigned alphanumeric user ID and summary usage statistics — to demonstrate the platform's utility to funders.

8 Features and limitations

What Autophon is: Autophon is a web-based application designed to simplify forced alignment workflows and expand access for users with minimal technical background. It is particularly useful for research on under-resourced languages and non-standard varieties, and emphasizes ease of use, format flexibility, and language model diversity. The backend relies on existing forced alignment technologies developed over the past decades, wrapped in a modern frontend that facilitates fast, OS-independent processing.

Key features include:

1. Fully web-based and platform-independent (OS-agnostic).

[9] https://fasealign.readthedocs.io/en/latest/development.html

[10] Users working near this cutoff time—e.g., at 2:55 AM GMT—should be aware that their files may disappear if alignment is not initiated in time.

2. No programming or installation required.
3. Accepts a wide range of transcription and audio formats.
4. Capable of processing low-resource and non-standard language varieties.
5. Supports user-defined pronunciation dictionaries and multiple transcription modes.

What Autophon is not: Important caveats to bear in mind:

1. Alignment quality depends on transcription accuracy, recording quality, and language characteristics.
2. Performance may vary across languages, dialects, and speaking styles.
3. Benchmarking accuracy is ongoing and not available for all models.
4. Core updates to underlying alignment engines may not be immediately reflected, due to the complexity of the Autophon backend.

9 Budget and funding

Autophon costs approximately 25 000 SEK (2 300 EUR) per year to run. Founded by Dr. Nate Young (who is the sole copyright holder), the project has since received support from the University of Helsinki, Linnaeus University, The Swedish Academy, the Department of Linguistics and Scandinavian Studies at the University of Oslo, and the European Union's *Horizon 2020* research and innovation programme under the *Marie Skłodowska-Curie* grant agreement No. 892963. Additional funding for language model development has come from The National Library of Norway[11].

We continue to seek funding and welcome collaboration. If you are experienced in grant writing or interested in supporting the project, please reach out via the support page.

Acknowledgements

Numerous individuals have contributed to Autophon. We especially thank Michael McGarrah for strategic guidance and Kaosi Anikwe for extensive backend and frontend development. Ismail Raji Damilola helped implement a bootstrapping function to expand phoneme inventories. Additional contributions in the early stages came from Nabil Al Nazi, Zamanat Abbas Naqvi, and Santiago Recoba.

References

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [softw.], ver.6.0.36. www.praat.org

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Proceedings of Interspeech*, 498–502.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., et al. (2011). *The Kaldi speech recognition toolkit* (tech. rep.). IEEE Signal Processing Society. Piscataway.

Rosenfelder, I., Fruehwald, J., Brickhouse, C., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2022). FAVE (Forced Alignment and Vowel Extraction) Program Suite v2.0.0 [Zenodo].

Wilbanks, E. (2015). The development of FASE (Forced Alignment System for Español) and implications for sociolinguistic research. *paper presented at New Ways of Analyzing Variation 44, University of Toronto*, 22–25 October. http://ericwilbanks.github.io/files/wilbanks_nwav_2015.pdf

Wilbanks, E. (2022). faseAlign (Version 1.1.14). https://github.com/EricWilbanks/faseAlign

Young, S. J., Woodland, P. C., & Byrne, W. J. (1993). *HTK: Hidden Markov Model Toolkit V1. 5*. Cambridge Univ. Eng. Dept. Speech Group; Entropic Research Lab. Inc.

[11] https://www.nb.no/sprakbanken/ressurskatalog/oai-nb-no-sbr-59/

Appendix

The Development of **FASE**: Forced Alignment System for Español and Implications for Sociolinguistic Methodologies

Eric Wilbanks

North Carolina State University

NWAV 44, Toronto
October 24, 2015

Overview of Talk

Motivation

Acoustic Models

Application

Motivation

Forced Alignment

- Over the past decade, technologies from speech recognition have begun to be utilized in phonetic research.
- Forced alignment takes as input an orthographic transcription and audio file and creates as output a time-aligned phonological (or possibly phonetic) transcription.

Benefits for Phonetics

- Manual segmentation of phones is incredibly time-consuming, at some estimates 800x real-time (Schiel and Draxler, 2003).
- Completely automated transcription/segmentation is still a work in progress (c.f. Reddy and Stanford, 2015)
- Automated segmentation, however, is increasing by orders of magnitude the amount of acoustic data linguists are able to analyze.
- As Labov et al. (2013) note, utilizing forced alignment allowed them to increase tokens extracted from each interview from 300 to 9,000.

P2FA

- The mostly widely used acoustic models used for English forced aligning are part of the **Penn Phonetics Lab Forced Aligner** (Yuan and Liberman, 2008, P2FA).
- Trained on a large corpus of Supreme Court Justice oral argument recordings; Extremely robust for North American English
- These acoustic models are also adapted for use in the **Forced Alignment and Vowel Extraction** suite (Rosenfelder et al., 2011, FAVE).

Non-English

- Comparable systems for languages other than English are not yet as widely researched or utilized.
- **Prosodylab Aligner** (Gorman et al., 2011) provides models for NA English and Quebec French and also supports training of novel models.
- **SPLaligner** (Milne, 2014) French aligner trained on Canadian political recordings
- **PraatAlign** (Lubbers and Torreira, 2015) Praat plugin with support for a variety of languages, including Spanish
- **EasyAlign** (Goldman, 2011) supports semi-automated alignment of various languages (including Spanish) from within Praat. Spanish models are trained on 2.9 hours of Castilian read speech.

Goals

- Report on the validity of utilizing messy sociolinguistic interviews to train forced alignment systems, in place of clean read speech.
- Argue for speaker adaptations by linear transforms in both training and alignment to improve alignment across a variety of recording environments and speakers.
- Demonstrate application of aligner: /d/ lenition within the corpus

Acoustic Models

Hidden Markov Models

- Hidden Markov Models (HMMs) take a sequence of observations (in our case acoustic vectors) and give them some label (phones)
- This is done by modeling each label/phone as a sequence of "hidden" states.
- During training, observations are paired with labels so that transition probabilities between states and model vectors can be learned.

Dictionary Construction

- In order to carry out training and aligning, a pronunciation dictionary is needed which maps words to strings of phones.
- The dictionary was constructed from the 44 million words **SUBTLEX-ESP** corpus (Cuetos et al., 2011).
- Spanish orthography is very close to phonological representation, making conversion of words to phone sequences easy.
- English loan words removed from corpus by cross-referencing with CMU Pronouncing Dictionary (Weide, 1994) and manually sorting.
- Final Spanish Pronunciation Dictionary - 93,350 unique words

Monophone Inventory

	labial	dental	alveolar	palatal	velar
plosives - voiceless	p	t			k
plosives - voiced	b	d			g
fricatives - voiceless	f		s		x
fricative - voiced				j (y)	
affricate				tʃ (CH)	
nasals	m		n	ɲ (NY)	
lateral			l		
rhotic - tap			ɾ (r)		
rhotic - trill			r (R)		

Vowels: /a,e,i,o,u/ correspond to their ipa symbols
Non-Speech: Laughing (lg), Coughing (cg), Breath (br), Noise (ns), short pause (sp), silence (sil)

Processes to Note

- Latin American Spanish, therefore we have no /s/-/θ/ or /l/ - /ʎ/ distinctions
- No distinction made between tonic and atonic vowels
- No distinction between high vowels and their glide allophones

Technical Specifications

Model training and application was carried out using the HTK suite (Young et al., 2006) with the following parameter values.

- WAV files downsampled to 11025hz, 11 mfcc coefficients, delta, and delta-delta extracted
- Emitting state of **sp** tied to the central state of the **silence** phone.

	States	Gaussians
sp	3	32
sil	5	32
others	5	16

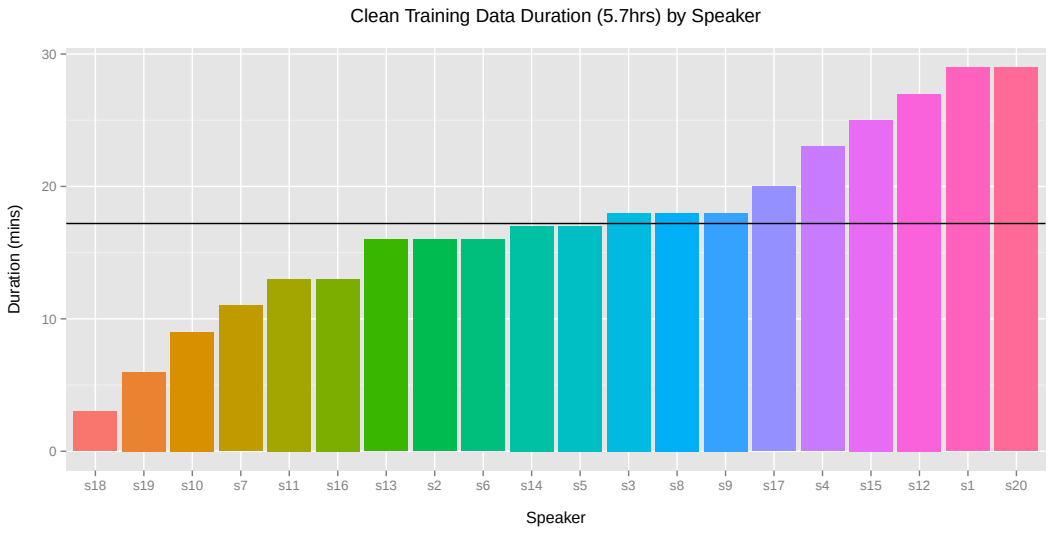
CERD

- The Corpus del Español de Raleigh-Durham (CERD) contains over 240 sociolinguistic interviews conducted in Spanish between 2008-present.
- Speakers come from a variety of regions, though most speakers (or their families) are from Mexico, Colombia, or Puerto Rico.
- Include variable experience with English, Heritage Speakers to 1-2 years in the US.

Transcription

- 20 Interviews were chosen to be orthographically transcribed and used as training data.
- Balanced for sex and age group
- Speakers were either from Mexico or of Mexican descent.
- Notably, the variety of Spanish spoken in central Mexico tends to resist elision processes typical of other varieties. Ideal for training models.
- Orthographic transcriptions were carried out by native Spanish L1 speakers

Training Data



Manual Segmentation

- Two trained Spanish phoneticians individually hand-segmented 100s of speech from a sociolinguistic interview.
- Speaker is external to the training data, young female speaker born in Mexico who moved to North Carolina at a young age.
- Differences between boundary placement/segment duration are computed between the two human transcribers and between each transcriber and the model.

Model Comparison

- **M1**: no speaker adaptation during training
- **M2**: adaptation during training; acoustic models updated via Constrained Maximum Likelihood Linear Regression (CMLLR) transforms

Why might we need Adaptation?

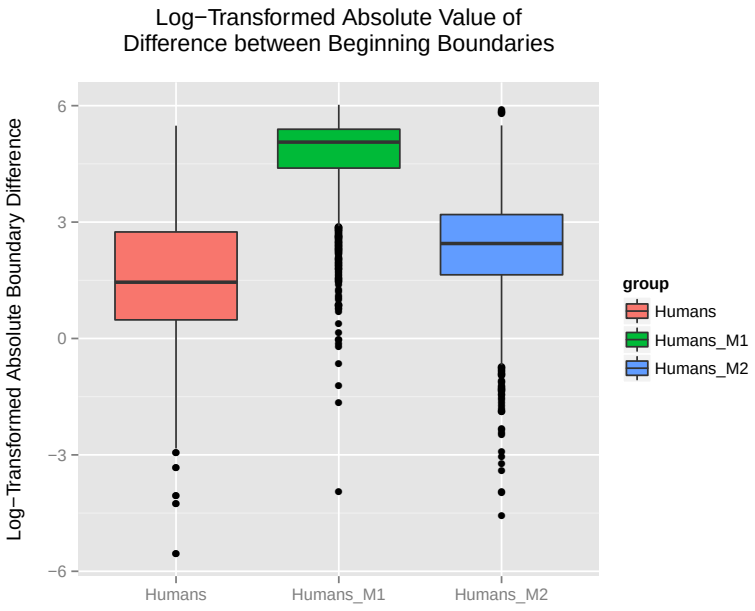
Good Quality

"El clima, ahorita está haciendo buen clima, pero sí cuando se cerca invierno,"
The weather, right now the weather's good, but yeah when winter comes,

Bad Quality

"Y su hermano menor estudia,"
And their younger brother studies,

Log Beginning



Beginning Linear Model

1. No sig. difference between HumanA-Model and HumanB-Model ($p = 0.17$)
2. **begin_diff** significantly lower in HumanA-HumanB group than in HumanA-Model and HumanB-Model ($p < 0.001$) groups.
3. M2 (adapted) has significantly lower boundary differences than Model 1; ($\beta = -3.616, p < 0.001$)

Linear model; dependent: absolute value of Beginning Difference, independent: Group and Model

lm(abs(begin_diff) ~ group + model, data = M1M2)

Distributions

	mean	sd	se
HumanA_HumanB	14.47ms	25.57	0.92
Humans_Model1	26.23ms	44.53	1.13
Humans_Model2	20.81ms	31.99	0.81

Descriptive Statistics of Boundary Differences by Group for Model2

Comparisons

	$<$ **10ms**	$<$ **20ms**
Goldman (2011)[1]	60.26%	87.11%
HumanA_HumanB	68.38%	78.66%
Humans_Model1	37.28%	63.30%
Humans_Model2	45.05%	69.41%

Percentage of Boundary Differences by Group for Models 1 and 2

[1] Human-Model; Read data rather than spontaneous

Application

/b,d,g/ Lenition - [β,ð,ɣ]

- Spanish voiced stops alternate between occlusive and approximant realizations; traditionally considered a binary distinction (Tomás, 1967)
- Recent work demonstrates it's best considered a gradient process (Lewis, 2001)
- Acoustic/Articulatory realizations conditioned by a variety of segmental, prosodic, lexical, and morphological variables

/d/ Specifically

Examining intervocalic /d/ we expect to see the most occlusion

1. After high vowels (Simonet et al., 2012)
2. Before high vowels (Ortega-LLebaría, 2004)

Additionally, preceding environment tends to have the stronger effect (Simonet et al., 2012).

Methodology

- All intervocalic /d/ tokens from the 20 training speakers extracted
- Exclusion of "De" (21% of data) leaves 1,482 tokens.
- Following Hualde et al. (2011), the difference in intensity between the following vowel and the /d/ is calculated.
- If a value is closer to 0, it indicates the /d/ is more open and less occlusive

Intensity Example

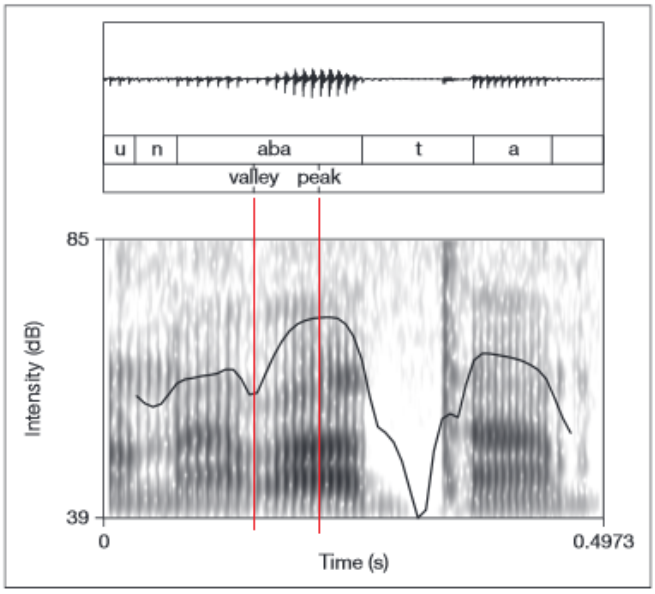
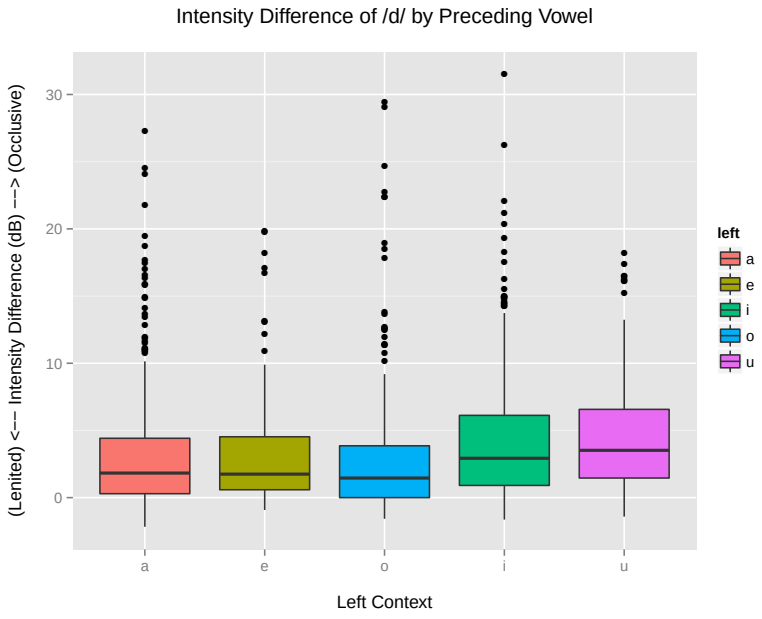
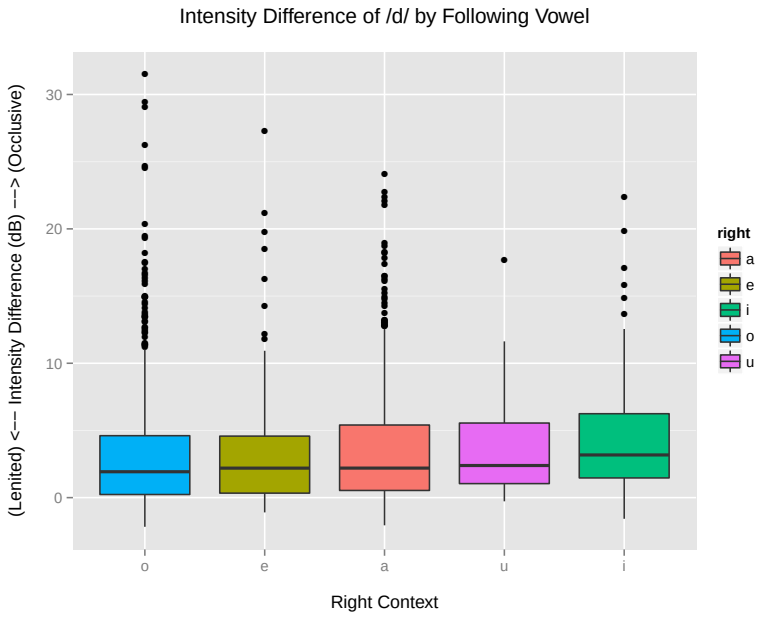


Image adapted from Carrasco et al. (2012, pp. 156)

Preceding Segment



Following Segment



Linear Mixed Model

- /d/ sig. more occlusive when preceded by /i,u/ > /a,e,o/
- /d/ sig. more occlusive when followed by /i/ > /u,a,e,o/

lmer(intensity_diff ~ left + right + (1|speaker),data = df)

Take Home Points

- Using speaker adaptation, sociolinguistic corpora make excellent training data for new forced aligners
- FASE produces excellent alignments of novel data, although not surpassing human transcription
- Using automatic alignments, well-studied internal constraints of /d/ lenition were reproduced within the corpus.

References

References

Carrasco, P., Hualde, J. I., and Simonet, M. (2012). Dialectal differences in spanish voiced obstruent allophony: Costa rica versus iberian spanish. *Phonetica*, 69:149–179.

Cuetos, F., Glez-Nosti, M., Barbón, A., and Brysbaert, M. (2011). Subtlex-esp: Spanish word frequencies based on film subtitles. *Psicológica*, 32:133–143.

Goldman, J.-P. (2011). Easyalign: an automatic phonetic alignment tool under praat. Proceedings of *Interspeech*, Firenze, Italy.

Gorman, K., Howell, J., and Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, 39(3):192–193.

Hualde, J. I., Shosted, R., and Scarpace, R. (2011). Acoustics and articulation of spanish /d/ spirantizaion. In *Proceedings of the 19th International Congress on the Phonetic Sciences, Hong Kong*.

Labov, W., Rosenfelder, I., and Fruehwald, J. (2013). One hundred years of sound change in philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 89(1):30–65.

Lewis, A. M. (2001). *Weakening of Intervocalic /P, T, K/ in Two Spanish Dialects: Toward the Quantification of Lenition Processes*. PhD thesis, University of Illinois at Urbana-Champaign.

Lubbers, M. and Torreira, F. (2013-2015). Praatalign: an interactive praat plug-in for performing phonetic forced alignment. `https://github.com/dopefishh/praatalign`. Version 1.7a.

Milne, P. (2014). *The variable pronunciations of word-final consonant clusters in a force aligned corpus of spoken French*. PhD thesis, University of Ottawa.

Ortega-LLebaría, M. (2004). Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: comparing intervocalic /b/ and intervocalic /g/ in spanish and english. In Face, T. L., editor, *Laboratory approaches to Spanish phonology*, pages 237–253. Mouton de Gruyter, Berlin, 1 edition.

Reddy, S. and Stanford, J. N. (2015). A web application for automated dialect analysis. Proceedings of NAACL 2015.

Rosenfelder, I., Fruehwald, J., Evanini, K., and Yuan, J. (2011). Fave (forced alignment and vowel extraction) program suite. http://fave.ling.upenn.edu.

Schiel, F. and Draxler, C. (2003). *The production of speech corpora*. Bavarian Archive for Speech Signals.

Simonet, M., Hualde, J. I., and Nadeu, M. (2012). Lenition of /d/ in spontaneous spanish and catalan. In *Interspeech*, pages 1416–1419.

Tomás, T. (1967). *Manual de pronunciación española*. Consejo Superior de Investigaciones Científicas. Instituto "Miguel de Cervantes." Publicaciones de la Revista de filología española. Graficas Monteverde, S.A.

Weide, R. L. (1994). Cmu pronouncing dictionary. `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`.

Young, S. J., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. C. (2006). *The HTK Book, version 3.4*. Cambridge University Engineering Department, Cambridge, UK.

Yuan, J. and Liberman, M. (2008). Speaker identification on the scotus corpus. Proceedings of Acoustics '08.

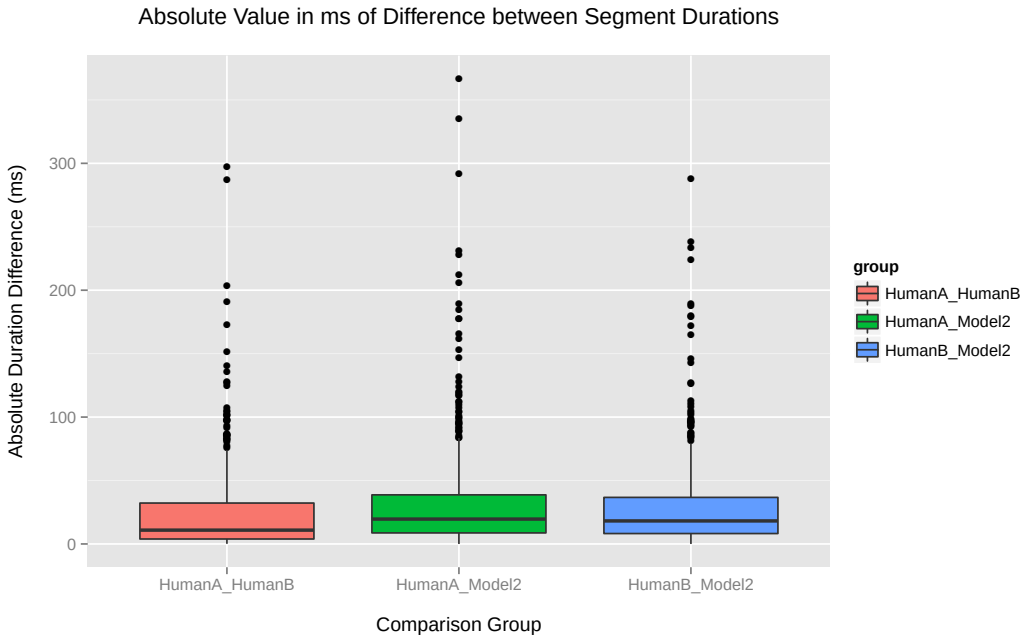
Thank you! Gracias!

Questions, Comments, Suggestions?

@eric_wilbanks
ewwilban@ncsu.edu
ericwilbanks.github.io

References

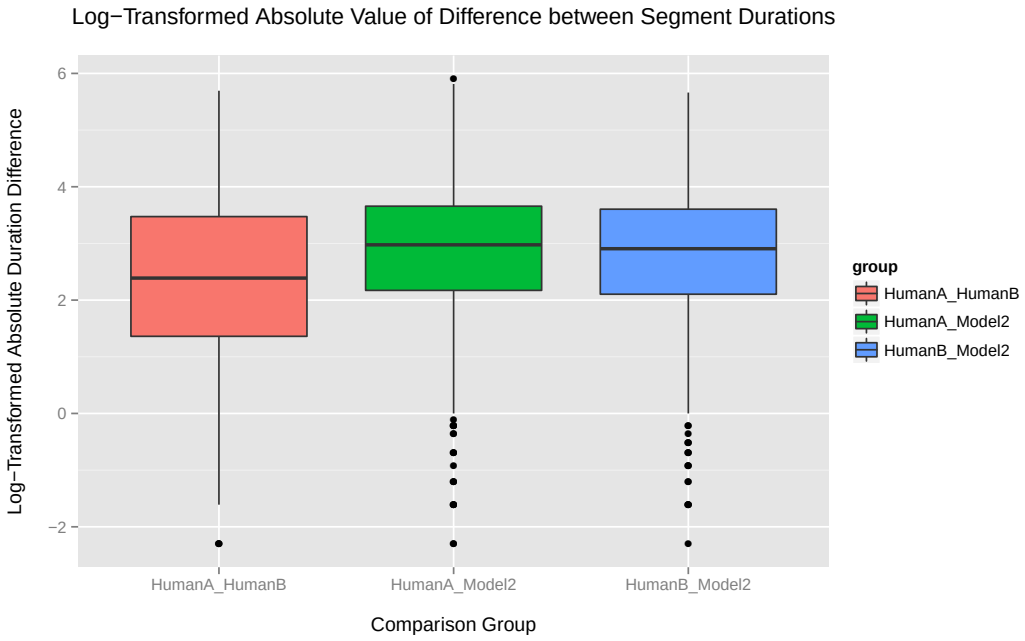
Duration



Eric Wilbanks NWAV 44

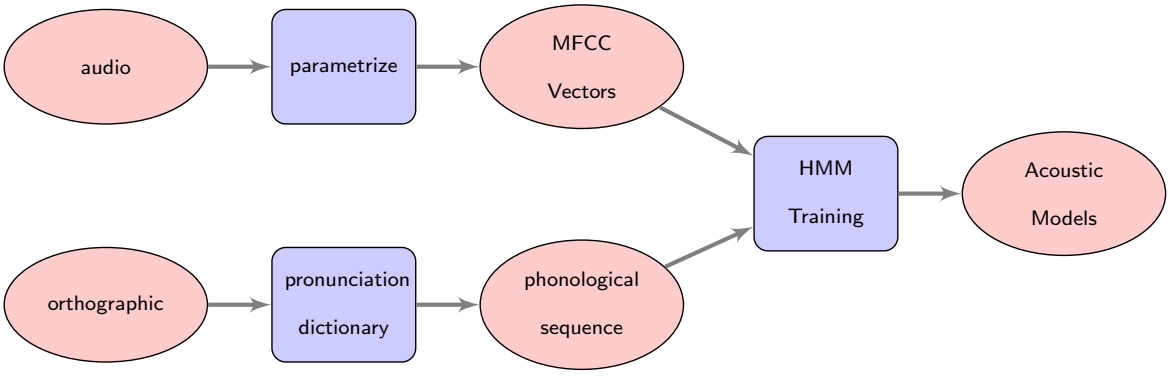
References

Log Duration



References

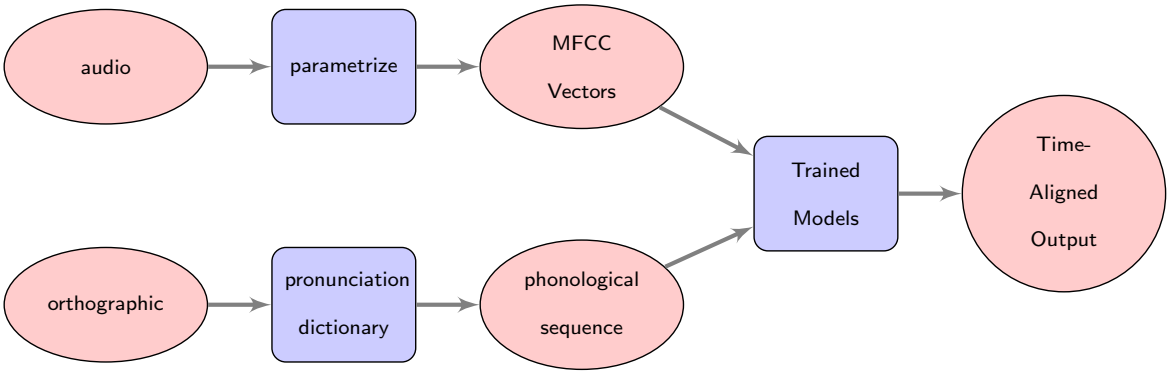
Training Acoustic Models



Eric Wilbanks NWAV 44

References

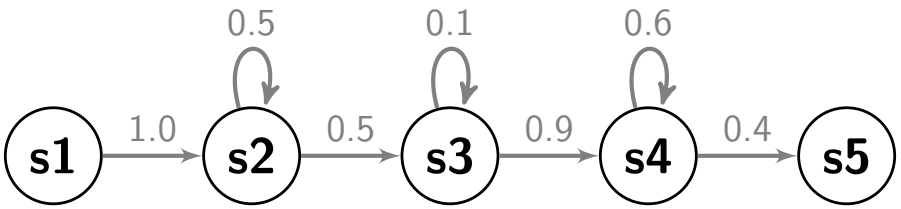
Generating Alignments



Eric Wilbanks NWAV 44

References

Left-To-Right HMM Model of Phone



Eric Wilbanks NWAV 44

References

Linear Mixed Model

Left	a	e	i	o	u
a	x	0.21	-3.53	1.05	-3.68
e	-0.21	x	-2.87	0.69	-3.47
i	3.53	2.87	x	3.76	-1.34
o	-1.05	-0.69	-3.76	x	-3.98
u	3.68	3.47	1.34	3.98	x

Right	a	e	i	o	u
a	x	-0.95	-2.44	-0.26	-0.61
e	0.95	x	-1.09	0.79	-0.16
i	2.44	1.09	x	2.23	0.42
o	0.26	-0.79	-2.23	x	-0.54
u	0.61	0.16	-0.42	0.54	x

Columns are Reference Levels

$$\mathbf{lmer(intensity\_diff \sim left + right + (1|speaker), data = df)}$$

References

Linear Mixed Model

Left	a	e	i	o	u		Right	a	e	i	o	u
a	x						a	x				
e		x					e		x			
i			x				i			x		
o				x			o				x	
u					x		u					x

Columns are Reference Levels
Green - Sig. More Lenited /d/
Red - Sig. More Occlusive /d/

lmer(intensity_diff ~ left + right + (1|speaker),data = df)