Autophon user guide
Icelandic

Engine: Montreal Forced Aligner 1.0
Model: IceFA 1.0

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 ***Icelandic***, using the ***Montreal Forced Aligner 1.0*** engine with the ***IceFA 1.0*** 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**.

Bootstrapping For this language, Autophon uses a *bootstrap* model. Instead of training an entirely new acoustic model for Icelandic, Autophon reuses an existing model originally developed for Norwegian. This is a common strategy in the development of speech tools for low-resource languages, where researchers often begin with models from more well-resourced languages before adapting them to new ones[5].

The bootstrapping process was carried out as follows:

1. The phonemes of both Icelandic and Norwegian were identified using the distinctive feature tables in Hayes (2009, pp. 95–97).
2. Levenshtein distances were calculated between the phonemes of the two languages, following methods such as those in Gooskens and Heeringa (2004), to determine phonemic similarity.
3. Each phoneme in Icelandic was matched with the closest phoneme in Norwegian based on the smallest Levenshtein distance.
4. In cases of ties (i.e., multiple equally close candidates), phoneme frequency was used to select the best match. This involved compiling a word frequency list[6], mapping those words to their phonemic representations, calculating phoneme frequencies, and selecting the most frequent candidate.

If your project requires the phoneme mappings, distance matrices, or frequency calculations, please contact tech support. We will be glad to provide the datasets by email.

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

[2] faseAlign was built by Wilbanks (2022). Like FAVE, it relies on the Hidden Markov Toolkit (S. J. 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).

[5] DiCanio, Nam, Whalen, Timothy Bunnell, Amith, and García (2013); Strunk, Schiel, Seifart, et al. (2014); Coto-Solano and Solórzano (2017); Coto-Solano, Nicholas, and Wray (2018); N. J. Young and McGarrah (2023)

[6] https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists

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**.

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[7]. 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:[8] 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.

[7] https://commonvoice.mozilla.org
[8] This field originally drove the development of forced alignment in the early 2000s.

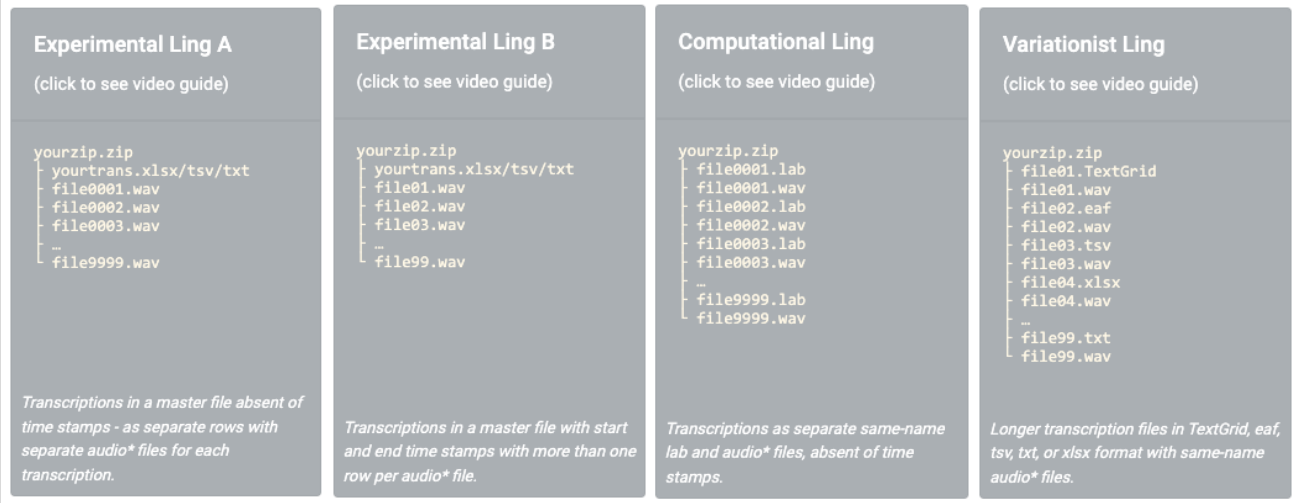


Figure 1: The Transcription Mode selection menu for Autophon.

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

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.[9]

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.

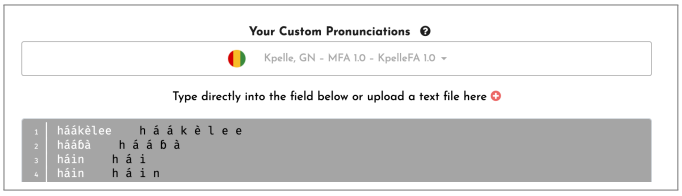
Pronunciations must be entered using the alphanumeric string specific to the language model at hand — in this case, the ***Grammatek X-Sampa***. Section 4 provides a key that maps the Grammatek X-Sampa to its IPA[10] equivalents.

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.

[9] 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.

[10] International Phonetic Alphabet

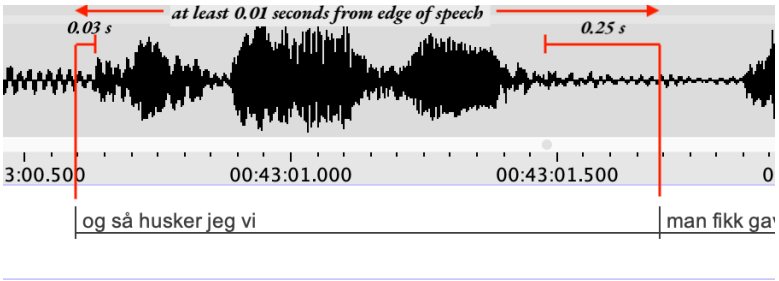


Figure 3: A sample transcription with at least 0.01 seconds of buffer on either end of the speech stream.

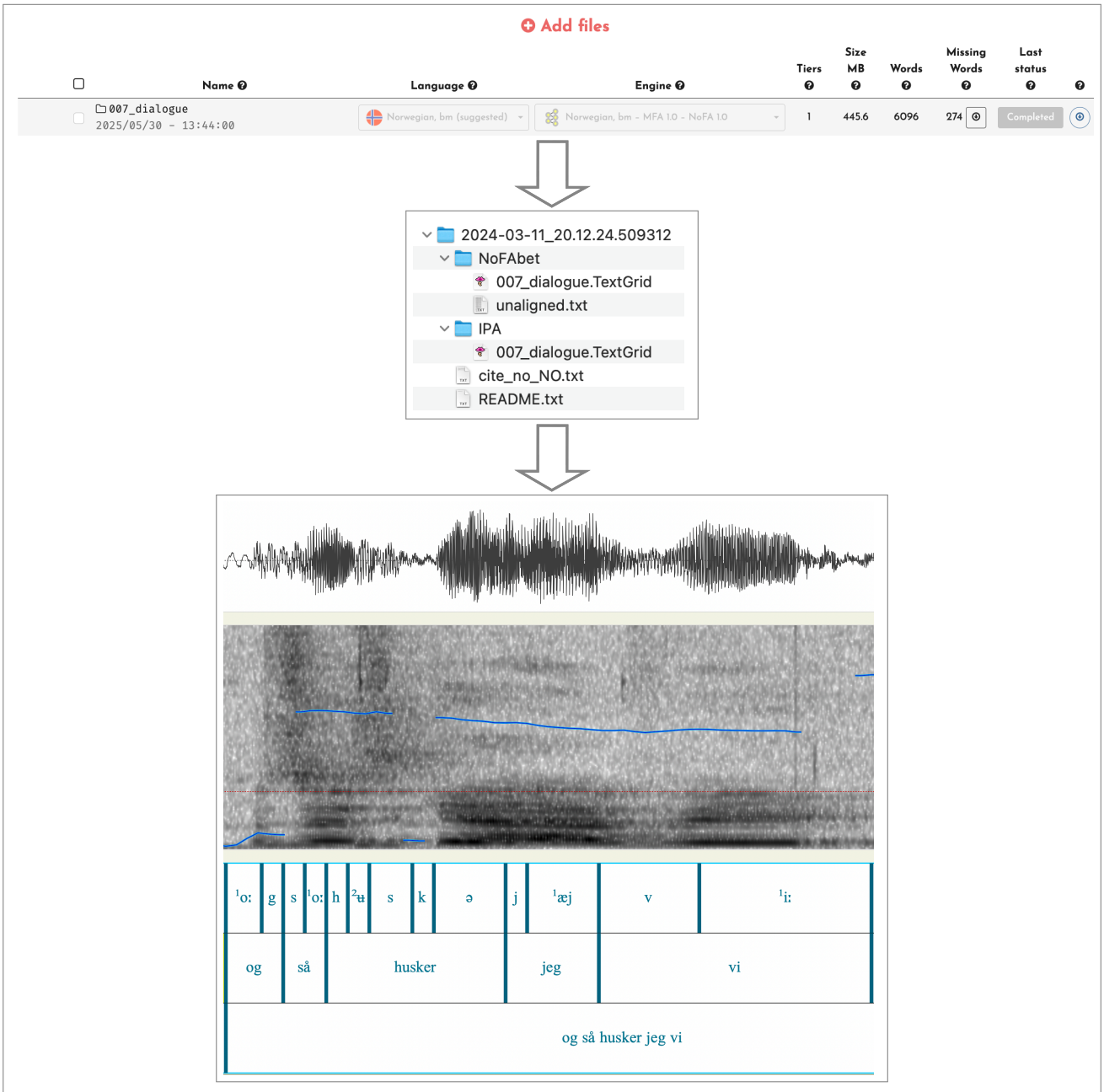


Figure 4: The alignment process, including task list, folder structure, and Praat TextGrid.

3 How to cite

Any dissemination or publication that makes use of this Autophon package for ***Icelandic***, which uses ***IceFA 1.0*** within ***The Montreal Forced Aligner 1.0*** 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

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.

Nikulásdóttir, A. B., Ármannsson, B., Bergþórsdóttir, B., & Rögnvaldsson, E. (2023, October). Icelandic Pronunciation Dictionary for Language Technology [Release 23.10]. https://github.com/grammatek/iceprondict

Young, N. J. (2024). IceFA 1.0 – Forced Alignment of Icelandic, ver. 1.0. www.autophon.org

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 Grammatek X-Sampa specific to IceFA 1.0 for The Montreal Forced Aligner 1.0 and (2) a TextGrid in the International Phonetic Alphabet (IPA). The phoneme key is located in Table 1.

X-Sampa	IPA	example	X-Sampa	IPA	example	X-Sampa	IPA	example	X-Sampa	IPA	example
Full vowels			*Diphthongs*			*Consonants*					
a	a	aska	ai	ai	ætla	c	c	gera	r	r	fara
a:	a:	aka	ai:	ai:	æfa	c_h	cʰ	kæla	r_0	r̥	hraði
i	i	íska	au	au	ást	f	f	finna	s	s	sofa
i:	i:	nýta	au:	au:	láta	h	h	halda	t	t	dalur
u	u	kú)stur	ei	ei	einn	j	j	játa	t_h	tʰ	tala
u:	u:	súpa	ei:	ei:	leika	k	k	galdur	v	v	vera
9	œ	öskur	0i	ɔi	bogi	k_h	kʰ	kaldur	x	x	rækta
9:	œ:	böl	ou	ou	ósk	l	l	sæla	C	ç	hjóla
0	ɔ	ostur	ou:	ou:	ljós	l_0	l̥	velta	D	ð	viður
0:	ɔ:	nota	9i	œy	haust	m	m	muna	G	ɣ	saga
E	ɛ	nesti	9i:	œy:	laus	m_0	m̥	glampa	N	ŋ	langur
E:	ɛ:	lesa	Yi	ʏi	hugi	n	n	næla	N_0	ŋ̊	hönk
I	ɪ	hitta				n_0	n̥	hnefi	J	ɲ	angi
I:	ɪ:	síður	*Stress*			p	p	bera	J_0	ɲ̊	banki
Y	ʏ	undra	%◌	ˈ◌	aska	p_h	pʰ	pera	T	θ	þunnur
Y:	ʏ:	muna									

Table 1: Grammatek X-Sampa, IPA, and lexical examples. The stress denotation means that `%` *is placed before a vowel or diphthong to denote primary stress. Unstressed vowels take no denotation.*

5 Phoneme key

The phoneme inventory for *Icelandic* is mostly identical to the inventory used for *Grammatek's Icelandic Pronunciation Dictionary for Language Technology*[11], albeit with some minor adjustments, namely the addition of primary stress

[11] Nikulásdóttir, Ármannsson, Bergþórsdóttir, and Rögnvaldsson (2023, October)

% in front of the first vowel for every word. We encourage users to inform us of errors and provide suggestions for changes.

6 Acoustic model and pronunciation dictionary

This specific Autophon package for ***Icelandic*** uses IceFA 1.0 within The Montreal Forced Aligner 1.0, which is a bootstrap of the acoustic model for Norwegian Bokmål, *NoFA 1.0*. The latter was trained on Norwegian Bokmål from the RUNDKAST[12] and NB Tale[13] corpora. The pronunciation dictionary is adapted from *Grammatek's Icelandic Pronunciation Dictionary for Language Technology*[14]. One adjustment we made was adding a stress denotation **%** to the first vowel in every word in the dictionary[15]. Note that this was done with a universal script, so exceptions to the left-strong rule will be incorrectly denoted here in version 1.0. Note also that secondary stress is also missing. We are looking for co-authors to help improve our model and dictionary for the next version, so kindly contact us if you are interested[16].

7 Performance metrics

We currently have no performance metrics for IceFA 1.0 within The Montreal Forced Aligner 1.0. If you are willing to provide us with a set of manually-corrected TextGrids in this language, we would be eager to validate our model with them (and update this document accordingly).

8 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[17].

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.

9 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).
2. No programming or installation required.
3. Accepts a wide range of transcription and audio formats.

[12] Amdal, Strand, Almberg, and Svendsen (2008)
[13] https://www.nb.no/sprakbanken/en/resource-catalogue/oai-nb-no-sbr-31/
[14] Nikulásdóttir, Ármannsson, Bergþórsdóttir, and Rögnvaldsson (2023, October)
[15] Árnason (2011, p. 271)
[16] Email n8.young@gmail.com
[17] 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.

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.

10 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[18].

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

Amdal, I., Strand, O. M., Almberg, J., & Svendsen, T. (2008). RUNDKAST: An Annotated Norwegian Broadcast News Speech Corpus. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. http://www.lrec-conf.org/proceedings/lrec2008/pdf/486_paper.pdf

Árnason, K. (2011). *The Phonology of Icelandic and Faroese*. Oxford University Press.

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

Coto-Solano, R., Nicholas, S. A., & Wray, S. (2018). Development of Natural Language Processing Tools for Cook Islands Māori. *Proceedings of the Australasian Language Technology Association Workshop 2018*, 26–33.

Coto-Solano, R., & Solórzano, S. F. (2017). Comparison of Two Forced Alignment Systems for Aligning Bribri Speech. *CLEI Electronic Journal*, 20(1), 2–1.

DiCanio, C., Nam, H., Whalen, D. H., Timothy Bunnell, H., Amith, J. D., & García, R. C. (2013). Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment. *The Journal of the Acoustical Society of America*, *134*(3), 2235–2246.

Gooskens, C., & Heeringa, W. (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change*, *16*, 189–207.

Hayes, B. (2009). *Introductory Phonology*. Wiley-Blackwell.

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.

Nikulásdóttir, A. B., Ármannsson, B., Bergþórsdóttir, B., & Rögnvaldsson, E. (2023, October). Icelandic Pronunciation Dictionary for Language Technology [Release 23.10]. https://github.com/grammatek/iceprondict

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].

Strunk, J., Schiel, F., Seifart, F., et al. (2014). Untrained Forced Alignment of Transcriptions and Audio for Language Documentation Corpora using WebMaUs. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation: May 26-31, 2014* (pp. 3940–3947). European Language Resources Association (ELRA).

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

Young, N. J., & McGarrah, M. (2023). Forced alignment for Nordic languages: Rapidly constructing a high-quality prototype. *Nordic Journal of Linguistics*, *46*(1), 105–131. doi.org/10.1017/S033258652100024X

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

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.