Autophon user guide
Kpelle – Guinea

Model: Montreal Forced Aligner 1.0 (French)

1 Introducing Autophon and forced alignment

Autophon is a **free** online forced aligner. *Forced alignment* (FA) refers to the automatic process by which speech recordings are phonetically time-stamped with the help of Hidden Markov models or Deep Neural Networks. Autophon uses the latter by means of the *Montreal Forced Aligner*[1], which is built on the Kaldi toolkit[2]. The app outputs a time-stamped phonetic annotation, readable in Praat (Boersma and Weenink 2017), that is based on an optimization of two user inputs: (1) the speech recording and (2) a corresponding orthographic transcription.

Forced alignment is important because it automates something that is resource-intensive when done manually. A typical phonetic annotation can take between 250 and 400 minutes per recorded minute. In a place like Scandinavia – where labor costs are high – this cost has presented a barrier for linguists.

For a forced alignment tool to work, an acoustic model must be trained on the specific language, and an accompanying pronunciation lexicon must be built that covers every word in the language (See section 5).

Numerous forced aligners are in circulation and available to download and use. However, they often are command-line based and rely on operating systems (OS) that may be outdated and/or incompatible with your OS. Therefore, *Autophon aims to offer an **OS-agnostic** and **user-friendly** option for phoneticians around the world.*

Bootstrapping For this language, Autophon has made use of a "bootstrap" model. This means that instead of training an entirely new model for *Kpelle*, we are using the model from *French* instead. It is relatively common practice for researchers of low-resource languages to bootstrap acoustic models from higher-resource languages before training a custom model[3]. The way we did this was as follows: (1) we first identified the properties of the phonemes in *French* and the phonemes in *Kpelle* according to the feature tables in Hayes (2009, pp. 95–97); (2) we calculated the Levenshtein distance between the phonemes as calculated, for example, in Gooskens and Heeringa (2004); (3) we matched the phonemes in *French* that had the shortest Levenshtein distance to phonemes in *Kpelle*; (4) for phonemes with more than one match, we consulted a word frequency list[4], mapped the words to their phonemic annotation, calculated phoneme frequency from this mapping, and then selected the phoneme with the highest frequency. If your project should require a detailed account of the phoneme maps, Levenshtein distance charts, and/or phoneme-frequency data, simply contact tech support, and we will email over the datasets.

2 Using the app

Aligning files without registering You can align smaller files immediately on the main page by clicking on *Add files* at the bottom. A box will appear with the heading *Transcription Mode: change transcription mode*. Click on the heading to select one of four *Transcription Modes* (see below). Once your transcription mode has been selected, select your files.

Registering and logging in For alignment of larger files and for access to the full suite, you must create a free account by clicking on *Sign up*. We require an account because we wish to keep track of usage in order to make a case for funders. Furthermore, an open system makes us vulnerable to bot attacks. Once you have registered, a verification email will be sent to you with a link that you must click on to verify your account. If you do not receive the email, first check your spambox and then wait at least 15 minutes before contacting tech support.

Cost Autophon is free of charge.

[1] McAuliffe, Socolof, Mihuc, Wagner, and Sonderegger (2017)

[2] Povey, Ghoshal, Boulianne, Burget, Glembek, Goel, Hannemann, Motlicek, Qian, Schwarz, et al. (2011)

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

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

Aligning files in a registered account Once you are registered and verified, go to the *Aligner* tab and click *Add files*. A box will appear with the heading *Transcription Mode: change transcription mode*. Click on the heading to select one of four *Transcription Modes*. Once your transcription mode has been selected, select your files.

Transcription modes Four different *transcription modes* are available, each named according to the field in which the format is most common: *Experimental Linguistics A*, *Experimental Linguistics B*, *Computational Linguistics*, and *Variationist Linguistics*. Each can be selected by clicking on one of the boxes illustrated in Figure 1. The boxes illustrate a typical file structure for each mode and provide a link to a video that offers detailed formatting instructions.

> **Experimental linguistics A:** In this mode, you upload a master transcription spreadsheet along with corresponding audio files – one by one or within a zip file. The master sheet should have two columns: column 1 holds the audio file names in your folder; column 2 holds the corresponding transcriptions. This format is similar to that used by, e.g., CommonVoice[5] and assumes that each audio file contains a short snippet of speech, which means that time stamps are *not* permitted. If you have a master transcription spreadsheet with time stamps, you are in the wrong transcription setting and need to select *Experimental Linguistics B*, described below. The master transcription sheet can be in a two-column Excel `xlsx` or tab-delimited file with either the extensions `txt` or `tsv`.
>
> **Experimental linguistics B:** In this mode, you also upload a master transcription spreadsheet with corresponding audio files– one by one or within a zip file. Unlike in mode A, it should have four columns: column 1 holds the names of the audio files in your folder; column 2 – start times; column 3 – end times; column 4 – transcription. This mode is designed for longer audio files that warrant multiple lines of transcription. The master transcription sheet should either be a four-column Excel `xlsx` or a tab-delimited file with either the extensions `txt` or `tsv`. Time stamps must be in *seconds formatted as real numbers*. European comma decimals are accepted (e.g., `1,23`) as well as Anglo-American period decimals (e.g., `1.23`). What will not work, however, are time stamps with colons, minutes, or hours (e.g., `00:00:01.23`).
>
> **Computational linguistics:** In this mode, you upload pairs of `lab` and audio files by the same name – one by one or within a zip file. These so-called `lab` files are simply text files that contain a single transcription phrase that matches the speech within the same-named audio file. Importantly, transcriptions should contain no time stamps. If you wish to have a complex set of subfolders within the zip file, as is common for comp-ling corpora like, e.g., NST[6], Autophon will output the finished TextGrids using the same folder structure. An example of one such folder hierarchy is shown in Figure 2
>
> **Variationist linguistics:**[7] In this mode, you upload pairs of transcription and audio files by the same – one by one or within a zip file. In contrast to the previous mode, transcriptions are longer and include time stamps that delineate the speech at the phrase level. Transcription files may be in Praat `TextGrid` or in ELAN `eaf` and may have multiple speaker tiers. Alternatively, transcription files may be in Excel `xlsx` or a tab-delimited `txt` or `tsv` file.[8]. You have the option of uploading a three-column or four-column file, depending on your needs. If the recording has multiple speakers, upload a four-column transcription file whereby column 1 holds the speaker name, column 2 – start time, column 3 – end time, and column 4 – transcription. If the recording has just one speaker, a four-column file is of course fine, but you may also upload a three-column file. Column 1 should hold the start time, column 2 – end time, and column 3 – transcription. Time stamps must be in *seconds formatted as real numbers*. European comma decimals are accepted (e.g., `1,23`) as well as Anglo-American period decimals (e.g., `1.23`). What will not work, however, are time stamps with colons, minutes, or hours (e.g., `00:00:01.23`).

Transcription codecs We have built Autophon so that it accepts transcription files in ***most*** codecs, and this is a vital feature for its OS-agnostic goal. Accepted codecs include, but are not limited to, UTF-8 Unix, UTF-16 Windows CRLF, Windows ISO Latin 1, and Windows ISO Latin 9. If you encounter errors, please email a sample file to tech support so that we can update our code with that format[9]

Audio codecs We have built Autophon so that it accepts audio files in ***most*** codecs, and this is a vital feature for its OS-agnostic goal. Accepted codecs include AAC(M4A), AC-3, AIFF, AIFF/24bit, AIFF/32bit, ALAC, FLAC, M4R, MP3, OGG, OPUS, WAV/8bit, WAV/24bit, WAV/32bit, WAV/A-law, WAV/mu-law, and WMA. Autophon will

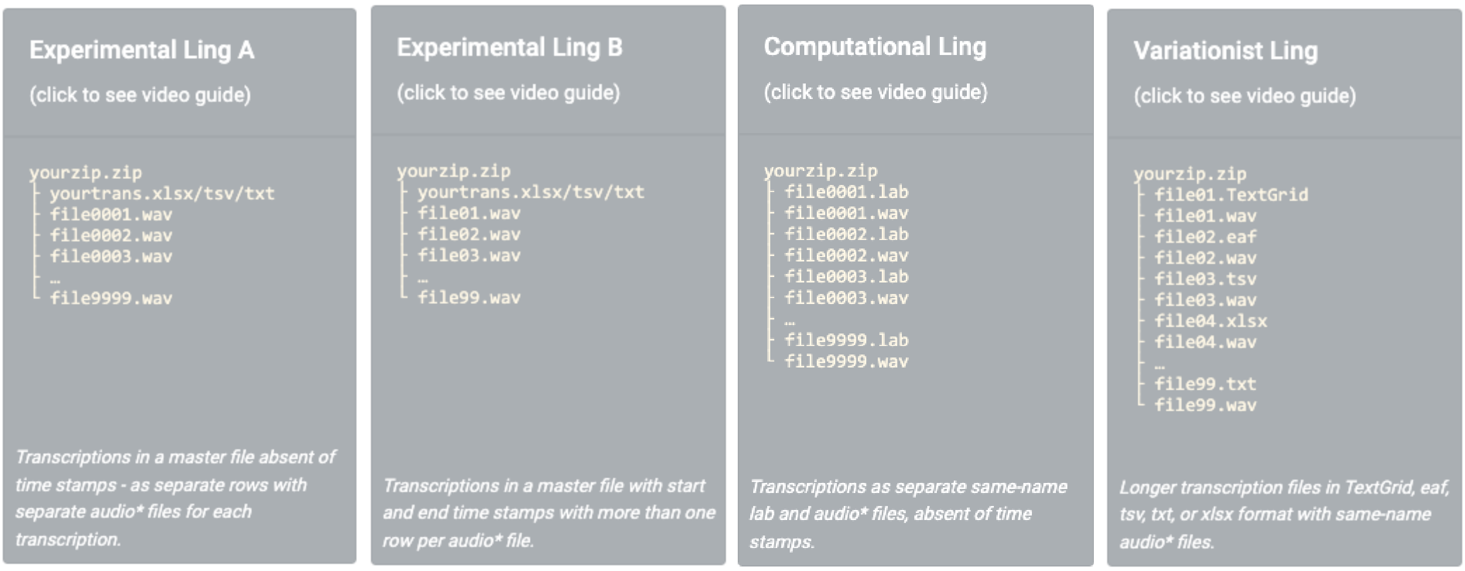
[5] https://commonvoice.mozilla.org

[6] https://www.nb.no/sprakbanken/en/resource-catalogue/oai-nb-no-sbr-16/

[7] This also happens to be the field that originally kickstarted forced alignment back in the early 2000s.

[8] The tab-delimited format is similar to the input format that was used for the legacy Penn Forced Aligner and FAVE Align.

[9] In the meantime, a quick fix is to open and resave them in a current version of Praat or ELAN.



Experimental Ling A

(click to see video guide)

```
yourzip.zip
├ yourtrans.xlsx/tsv/txt
├ file0001.wav
├ file0002.wav
├ file0003.wav
│ …
└ file9999.wav
```

Transcriptions in a master file absent of time stamps - as separate rows with separate audio files for each transcription.*

Experimental Ling B

(click to see video guide)

```
yourzip.zip
├ yourtrans.xlsx/tsv/txt
├ file01.wav
├ file02.wav
├ file03.wav
│ …
└ file99.wav
```

Transcriptions in a master file with start and end time stamps with more than one row per audio file.*

Computational Ling

(click to see video guide)

```
yourzip.zip
├ file0001.lab
├ file0001.wav
├ file0002.lab
├ file0002.wav
├ file0003.lab
├ file0003.wav
│ …
├ file9999.lab
└ file9999.wav
```

Transcriptions as separate same-name lab and audio files, absent of time stamps.*

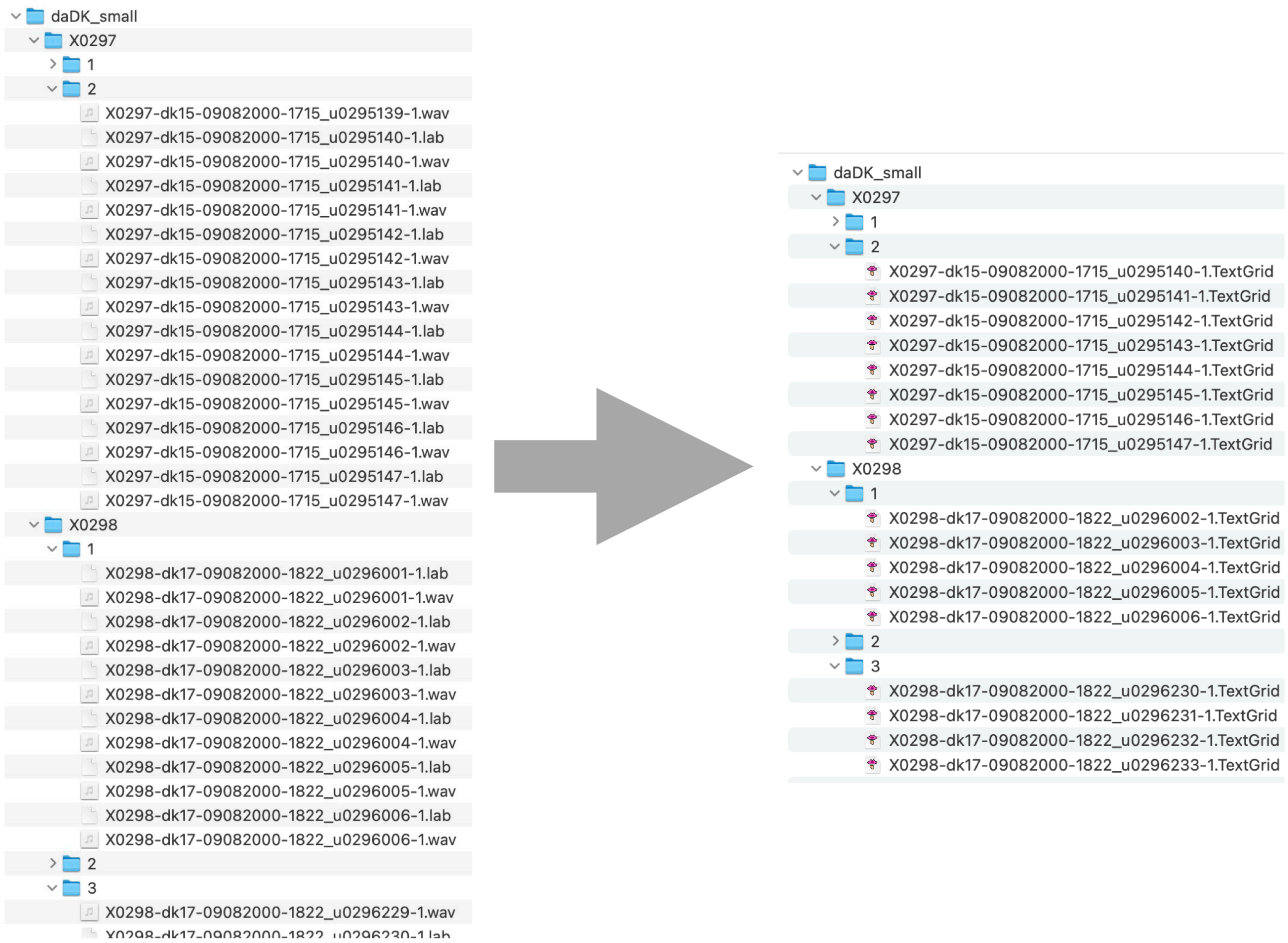
Variationist Ling

(click to see video guide)

```
yourzip.zip
├ file01.TextGrid
├ file01.wav
├ file02.eaf
├ file02.wav
├ file03.tsv
├ file03.wav
├ file04.xlsx
├ file04.wav
│ …
├ file99.txt
└ file99.wav
```

Longer transcription files in TextGrid, eaf, tsv, txt, or xlsx format with same-name audio files.*

Figure 1: The Transcription Mode selection menu for Autophon.



```
daDK_small
  X0297
    1
    2
      X0297-dk15-09082000-1715_u0295139-1.wav
      X0297-dk15-09082000-1715_u0295140-1.lab
      X0297-dk15-09082000-1715_u0295140-1.wav
      X0297-dk15-09082000-1715_u0295141-1.lab
      X0297-dk15-09082000-1715_u0295141-1.wav
      X0297-dk15-09082000-1715_u0295142-1.lab
      X0297-dk15-09082000-1715_u0295142-1.wav
      X0297-dk15-09082000-1715_u0295143-1.lab
      X0297-dk15-09082000-1715_u0295143-1.wav
      X0297-dk15-09082000-1715_u0295144-1.lab
      X0297-dk15-09082000-1715_u0295144-1.wav
      X0297-dk15-09082000-1715_u0295145-1.lab
      X0297-dk15-09082000-1715_u0295145-1.wav
      X0297-dk15-09082000-1715_u0295146-1.lab
      X0297-dk15-09082000-1715_u0295146-1.wav
      X0297-dk15-09082000-1715_u0295147-1.lab
      X0297-dk15-09082000-1715_u0295147-1.wav
  X0298
    1
      X0298-dk17-09082000-1822_u0296001-1.lab
      X0298-dk17-09082000-1822_u0296001-1.wav
      X0298-dk17-09082000-1822_u0296002-1.lab
      X0298-dk17-09082000-1822_u0296002-1.wav
      X0298-dk17-09082000-1822_u0296003-1.lab
      X0298-dk17-09082000-1822_u0296003-1.wav
      X0298-dk17-09082000-1822_u0296004-1.lab
      X0298-dk17-09082000-1822_u0296004-1.wav
      X0298-dk17-09082000-1822_u0296005-1.lab
      X0298-dk17-09082000-1822_u0296005-1.wav
      X0298-dk17-09082000-1822_u0296006-1.lab
      X0298-dk17-09082000-1822_u0296006-1.wav
    2
    3
      X0298-dk17-09082000-1822_u0296229-1.wav
      X0298-dk17-09082000-1822_u0296230-1.lab
```

```
daDK_small
  X0297
    1
    2
      X0297-dk15-09082000-1715_u0295140-1.TextGrid
      X0297-dk15-09082000-1715_u0295141-1.TextGrid
      X0297-dk15-09082000-1715_u0295142-1.TextGrid
      X0297-dk15-09082000-1715_u0295143-1.TextGrid
      X0297-dk15-09082000-1715_u0295144-1.TextGrid
      X0297-dk15-09082000-1715_u0295145-1.TextGrid
      X0297-dk15-09082000-1715_u0295146-1.TextGrid
      X0297-dk15-09082000-1715_u0295147-1.TextGrid
  X0298
    1
      X0298-dk17-09082000-1822_u0296002-1.TextGrid
      X0298-dk17-09082000-1822_u0296003-1.TextGrid
      X0298-dk17-09082000-1822_u0296004-1.TextGrid
      X0298-dk17-09082000-1822_u0296005-1.TextGrid
      X0298-dk17-09082000-1822_u0296006-1.TextGrid
    2
    3
      X0298-dk17-09082000-1822_u0296230-1.TextGrid
      X0298-dk17-09082000-1822_u0296231-1.TextGrid
      X0298-dk17-09082000-1822_u0296232-1.TextGrid
      X0298-dk17-09082000-1822_u0296233-1.TextGrid
```

Figure 2: Autophon will output the finished TextGrids using an identical subfolder structure as the uploaded file.

automatically consolidate stereo files to mono, *which may compromise quality due to phase cancellation*[10]. Therefore, you may wish to explicitly select either the left or right channel of your stereo file before aligning. If you encounter errors, please email a sample file to tech support so that we can update our code with that format[11]

Transcription preparation Regardless of what transcription mode you use, transcriptions should contain between one and 20 words. Boundary demarcations should have at least 0.01 seconds of buffer before and after the speech stream. This is illustrated in Figure 3, which shows a five-word phrase with a start boundary 0.03 seconds from the speech and an end boundary 0.25 seconds from the speech. Varying the boundary demarcation in this way is expected, and Autophon handles it well[12]

Select a language Once you upload your files into the aligner, it will suggest a language and language model. You are welcome to change the selection using the dropdown menu.

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

Missing words This feature can be understood if you have a basic understanding of how forced alignment works. Forced alignment maps a pre-defined phonemic pronunciation onto the speech stream by means of Deep Neural Networks. These pronunciations are defined by language-specific dictionaries that hold a finite list of words. The *missing words* feature provides a list of words not found in Autophon's dictionary and suggests a corresponding pronunciation. Autophon will simply default to using those suggestions for alignment, but you also can reject a suggestion and enter your own pronunciation. This process is described in the next section.

Your custom pronunciations As described above, forced alignment maps pre-defined phonemic pronunciations onto the speech stream by using language-specific dictionaries that hold a finite list of words. For missing words, Autophon suggests a pronunciation. You may decide that you either (a) do not agree with Autophon's missing words suggestions or that you (b) do not agree with the pronunciations within the language-specific dictionary. In this box you can enter your own pronunciations that will override both.

Pronunciations must be entered using the ASCII string specific to the language model at hand — in this case, African IPA. Table **??** holds a key for African IPA and its respective IPA[13] equivalents. You may type pronunciations directly into the dictionary box or upload them from a `txt` file. You are limited to 1 million characters. Entries must be formatted as `word-space-phoneme-space-phoneme` or `word-tab-phoneme-space-phoneme`, which means that each phoneme must be separated by a space. Note also that the lookup cannot be two or more words because that will confuse Autophon and make it treat the second word as a phone.

You may enter more than one pronunciation for the same word by repeating the word on the next line and providing a different pronunciation. Autophon will respond by attempting to find the most suitable pronunciation for that specific speech event. See below for examples of correct versus incorrect entries.

§ *Correct vs. incorrect entries in the "Your Custom Pronunciations" box.*

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)
```

[10] For more on phase cancellation, check out https://youtu.be/wY9QokRPJts

[11] In the meantime, a quick fix is to convert the file to WAV using software like FFmpeg or MediaHuman audio converter.

[12] If you have transcriptions of single words or phrases that are segmented at the *exact* start and finish times. Autophon will perform poorly and move those boundaries. This, however, is something we would be interesting in remedying by means of a fifth transcription mode, so kindly reach out to tech support if you have such a project.

[13] International Phonetic Alphabet

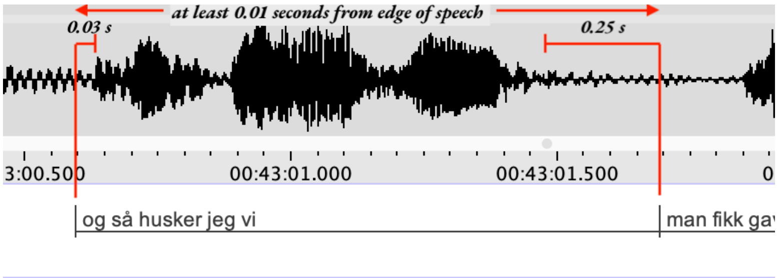
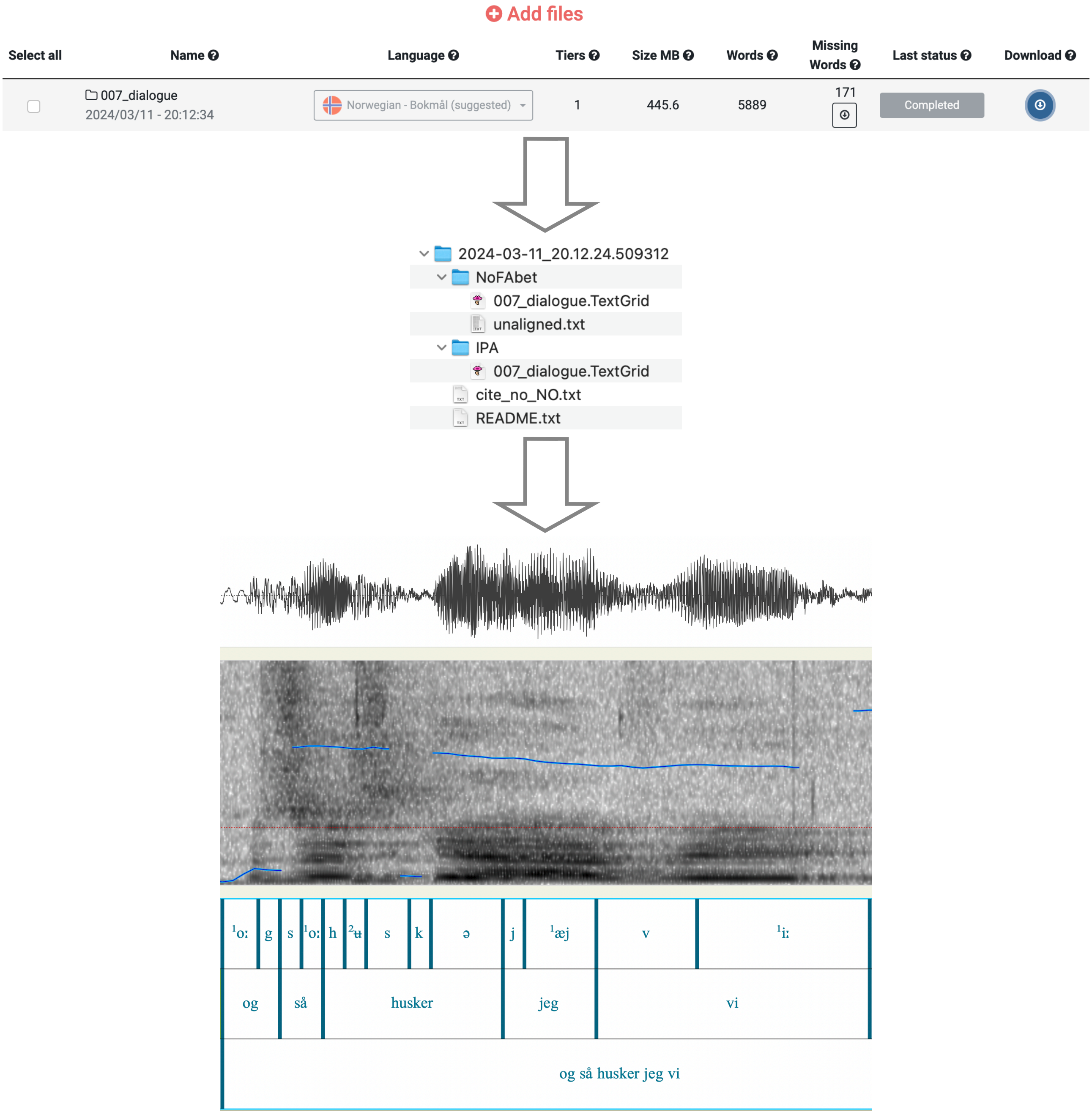


Figure 3: A sample transcription with at least 0.01 seconds of buffer on either end of the speech stream.



Select all	Name	Language	Tiers	Size MB	Words	Missing Words	Last status	Download
☐	007_dialogue 2024/03/11 - 20:12:34	Norwegian - Bokmål (suggested)	1	445.6	5889	171	Completed	

```
2024-03-11_20.12.24.509312
  NoFAbet
    007_dialogue.TextGrid
    unaligned.txt
  IPA
    007_dialogue.TextGrid
  cite_no_NO.txt
  README.txt
```

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

Aligning files Click *Align* to the far right of the upload list to initiate alignment. This will usually just take a few minutes, depending on how many people are using the aligner at that moment.

Downloading the annotations When alignment is finished, your annotations can be downloaded as Praat TextGrids via the downward arrow to the right of the task list. Figure 4 shows an example of this process.

3 How to cite

Any dissemination that makes use of Autophon *Kpelle – Guinea* should cite the below references. We understand that publishers often pressure researchers to slim down bibliographies; however, this should be done on the literature review and not on core analysis tools. Our view is plain: failure to cite constitutes plagiarism.

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.

Young, N. J., & Anikwe, K. (2024). Autophon – Automatic phonetic annotation of Nordic languages (web app). www.autophon.se

Young, N. J., & Konoshenko, M. (2024). KpelleFA 1.0 – Forced Alignment of Kpelle, ver. 1.0. www.autophon.org

4 Phoneme key

Not yet available.

5 Acoustic model and pronunciation dictionary

Kpelle – Guinea uses the very same acoustic model developed for version 1.0 of the Montreal Forced Aligner, which was trained on the *GlobalPhone* corpus[14]. The pronunciation dictionary we use was created by Maria Konoshenko. We encourage you to inform us of errors and provide suggestions for changes.

6 Performance metrics

Metrics for this specific model are not available.

7 Data security and G.D.P.R.

The files you upload to Autophon are encrypted and sent to a server in Frankfurt, Germany, that is run by *Digital Ocean*. Transcriptions and audio files are deleted immediately after alignment, which significantly reduces the chance of a data breach and keeps our costs low[15]. On the other hand, finished TextGrids are stored in your account for as long as you like. Once, however, you delete them, they will be removed from our server permanently.

If you upload any files and fail to click on *Align*, Autophon will delete them at 3AM Greenwich Mean Time[16].

We recognize our obligations to the European Union General Data Protection Regulation (GDPR), which is why we only collect four types of information from you: name, title, affiliation, and email address. Once you align a file, we permanently delete the audio. Once you delete the file from your task list, we also permanently remove the transcription and documentation of its original name. You may delete your account at any time, at which point we permanently delete your name, title, affiliation, and email address from our server. What we do *permanently* keep, however, is your alphanumeric account ID and the alignment activity linked to that ID — absent of original file names. We keep these records to show funders that Autophon is worth funding.

[14] https://montreal-forced-aligner.readthedocs.io/en/v1.0/pretrained_models.html

[15] We pay Digital Ocean approximately 90 USD per month for 60 GB of space, which means we have thin margins and cannot store much data. This also happens to keep Autophon's carbon footprint relatively low.

[16] Note that this means that if you are working late at night at, for example, 2.55 AM GMT, your uploaded files may disappear before you manage to align them. Bear this in mind.

8 Features and limitations

What Autophon is: Autophon is a frontend web application for the Nordic languages that uses the Montreal Forced Aligner (MFA) [17] as a core component of its backend. The language-specific models and pronunciation dictionaries were constructed by Dr. Nate Young. The most significant pieces of the app's backend were constructed by Kaosi Anikwe who joined the project in early 2023. The language-specific models are trained on various corpora, and the pronunciation dictionaries are usually adaptations of existing dictionaries available online.

The main advantages of using Autophon are:

1. Autophon is a web app, which means it is OS-agnostic.
2. As a web app, it requires no programming knowledge, which expands access to researchers and students.
3. Autophon accepts nearly all types of transcription and sound formats.
4. Autophon has a limitless repertoire of pronunciations by making use of grapheme-to-phoneme algorithms.
5. Autophon has models for Nordic languages, which have typically been neglected by forced alignment tech.

What Autophon is not: Important limitations are:

1. This is no magic bullet. Even with an accurate orthographic transcription, results may not satisfy.
2. Autophon varies in accuracy, and this accuracy depends on the language, speaker, and style.
3. Accuracy metrics are complex projects unto themselves, so they are unavailable for most languages.
4. Autophon will be slower to implement core MFA updates because it consists of layers and layers of code packed around MFA. For example, MFA 2.0 and 3.0 are not part of its backend yet.

9 Budget and funding

Autophon has cost ca. SEK 768 000 (ca. EUR 69 000) to develop and maintain since 2021. Initially started with private means by Dr. Nate Young, it has since grown with a grant from the Swedish Academy, a grant from the Department of Linguistics and Scandinavian Studies at The University of Oslo, and it has received funding from the European Union's *Horizon 2020 research and innovation programme* under the *Marie Sklodowska-Curie* grant agreement No 892963. Furthermore, The National Library of Norway funded development of the Norwegian Bokmål model[18].

We are actively looking for funders and collaborators who will support Autophon. We are also willing to share authorship with someone who can prepare grant applications and successfully procure funding. Contact us on the support page if you are interested.

Acknowledgements

Numerous individuals helped make Autophon possible. Michael McGarrah has offered important industry guidance, and Kaosi Anikwe has been an invaluable backend and frontend developer. Ismail Raji Damilola helped develop a bootstrapping function to adapt monophone inventories. The following programmers also worked on the app in its infancy: 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

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.

[17] McAuliffe, Socolof, Mihuc, Wagner, and Sonderegger (2017)

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

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.

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).

Young, N. J., & Anikwe, K. (2024). Autophon – Automatic phonetic annotation of Nordic languages (web app). www.autophon.se

Young, N. J., & Konoshenko, M. (2024). KpelleFA 1.0 – Forced Alignment of Kpelle, ver. 1.0. www.autophon.org

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