# A GUIDE TO CLOx

## Client Libraries Oxford


**Alicia Wassink, Rob Squizzero, Campion Fellin and David Nichols**

University of Washington

https://clox.ling.washington.edu/

# TABLE OF CONTENTS

# WHAT IS CLOx?

CLOx is a web-based service developed by the Sociolinguistics Laboratory at the University of Washington. CLOx uses Microsoft's Speech Service API **C**lient **L**ibraries, previously known as Project **Ox**ford, to orthographically transcribe sociolinguistic or other audio-recorded interviews in a format amenable to linguistic analysis. It outputs transcriptions in a .csv format with timestamps indicating the start and end time of each turn of speech contained in an audiofile. It is estimated that CLOx enables accurate transcription of a sociolinguistic interview to be completed in one-fifth or less of the time it would take to produce a manual transcription.

Currently supported languages and dialects include are represented in the table below. In the event that more languages are added to the Microsoft Speech Service API in the future, CLOx should be able to integrate them.

| Language | Region options |
|---|---|
| Arabic | Algeria, Bahrain, Egypt, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestinian Authority, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates, Yemen |
| Bulgarian | Bulgaria |
| Catalan | Spain |

| Language | Region options |
|---|---|
| Chinese | Cantonese traditional, Mandarin simplified, Taiwanese Mandarin |
| Croatian | Croatia |
| Danish | Denmark |
| Dutch | Netherlands |
| English | Australia, Canada, Ghana, Hong Kong, India, Ireland, Kenya, New Zealand, Nigeria, Philippines, Singapore, South Africa, Tanzania, United Kingdom, United States |
| Estonian | Estonia |
| Filipino | Philippines |
| Finnish | Finland |
| French | Canada, France, Switzerland |
| German | Austria, Germany |
| Greek | Greece |
| Gujarati | India |
| Hebrew | Israel |
| Hindi | India |
| Hungarian | Hungary |
| Indonesian | Indonesia |
| Irish | Ireland |
| Italian | Italy |
| Japanese | Japan |
| Korean | Korea |
| Latvian | Latvia |
| Lithuanian | Lithuania |
| Malay | Malaysia |
| Maltese | Malta |
| Marathi | India |
| Norwegian | Norway (Bokmål) |
| Polish | Poland |
| Portuguese | Brazil, Portugal |
| Romanian | Romania |
| Russian | Russia |
| Slovak | Slovakia |
| Slovenian | Slovenia |
| Spanish | Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Spain, Uruguay, USA, Venezuela |
| Swedish | Sweden |

| | |
|---|---|
| Tamil | India |
| Telugu | India |
| Thai | Thailand |
| Turkish | Turkey |
| Vietnamese | Vietnam |

CLOx has been tested with the Chrome, Firefox, Safari, and Opera web browsers. We do not recommend using CLOx without high speed internet, as slow connections are susceptible to timeout errors.

## CITING CLOX

You can cite CLOx as either of the following:

Wassink, A.B., Squizzero, R., Fellin, C. & Nichols, D. (2018). *Client Libraries Oxford (CLOx): automated transcription for sociolinguistic interviews* [computer software]. https://clox.ling.washington.edu.

Wassink, A.B., Squizzero, R., Fellin, C. & Nichols, D. (2018, October). *CLOx: Time-saving automated orthographic transcription tool for sociolinguistic interviews*. Workshop given at New Ways of Analyzing Variation 47, New York, NY.

# HOW DO I USE CLOx?

## REQUIRED SUBSCRIPTION KEY

CLOx works by directing your audiofile to Microsoft's Speech-to-Text server to be processed. To access this server, users need to have an access code called a "subscription key."

A free[1] subscription key may be acquired at http://azure.microsoft.com/. Consulting Microsoft's guide illustrating how to obtain your key to their speech service is highly recommended. Note that you will need to enter a credit card, even for a free account. Exception: students in 140 countries/regions may be eligible for a free account that does not require a credit card. See https://azure.microsoft.com/en-us/free/students/ for more information.

[1] The subscription is free for the first month, and after that provides up to 5 hours of transcription per month for free, then is billed at $1 per additional hour of transcription. Azure allows usage monitoring. For a very large amount of transcriptions done in a short time, you may incur fees. We are not responsible for any fees incurred using the service.

## PREPARING YOUR AUDIOFILE

File requirements at-a-glance:

- *.wav format*
- *mono/stereo: mono only*
- *maximum filesize: 19.2 MB*
- *sampling rate: 16 kHz*

CLOx operates on .wav files. Files uploaded must be mono (using one channel only), must be sampled at 16 kHz, and must be under 19.2 MB in size, which is 10 minutes of speech under those specifications, assuming 16-bit depth. If your file is not formatted in this way, you may (1) use one of our automated pre-processing options below to create a set of extracted files that meet these Microsoft Speech requirements, or (2) pre-process it manually.
**Warning:** if audio is not preprocessed, CLOx may crash without notification, produce undesirable output, or run at slower than optimal speeds.

### Automated pre-processing

We offer two methods for automated preprocessing: (1) a Praat script, (2) a Python script for Windows.

A Praat script, clox_preprocessing.praat, is provided with the CLOx service. This script will automatically format audiofiles to the CLOx requirements above. For example, many sociolinguists record audio in stereo at a 44.1 kHz sampling rate. Such files cannot be processed by the Speech Service[2]. This script does all the work necessary to extract

[2] Technically speaking, CLOx (and Microsoft Speech) will accept 44.1 kHz stereo audio, but Speech will automatically downsample audio to 16 kHz mono. This will cause CLOx to run more slowly without any accompanying improvement in output

one channel, trim silent intervals, lower the sampling rate, and generate temporary audiofiles of an appropriate length to work with the Speech Service. To obtain the script, navigate your browser to the CLOx website and click on the Scripts page. Launch Praat, then Open the script (select Praat, then Open Praat script...), and run it. You will be asked to specify a directory to save the preprocessed files in and to choose the sound to preprocess. A set of temporary audiofile extracts, indexed sequentially and meeting the CLOx requirements will be saved to your computer for use with CLOx.

CLOx Preprocessing is a Python-based command line program designed to convert audio files to be used with CLOx. The program takes a wav file of arbitrary length, converts it to mono (if necessary), and downsamples it to 16 kHz. If the file is larger than 10 MB, it breaks it into shorter files using silence detection to prevent interruptions of speech events. The output files append the start time in milliseconds as required by CLOx. The documentation for the Python script for Windows is available here.

## Manual pre-processing

To prepare your audiofile yourself, you may need to convert it from stereo to mono, resample the audio signal, and segment the file into smaller extracts. The steps below describe how to do this. These steps assume you are working in Praat, but any other signal analysis software program may be used (e.g., Audacity, etc.).

### *Step A: Converting from Stereo to Mono*

To extract one channel in Praat:

1. *Open your audiofile as a "Sound" object (not as a "LongSound").*

---

quality. In order to avoid timeout errors, CLOx will not accept files larger than 19.2 MB.

2. *Select the file in the object list.*
3. *Select "Convert -" then "Extract one channel...” and select the channel you wish to extract.*
4. *Save your new audiofile (in .wav format).*
5. *If your new (mono) file exceeds 10 minutes, either submit it to the preprocessor as in step 1 above for segmenting and resampling (recommended) or follow the steps below for Resampling (B) and Segmenting (C).*
6. *If your new (mono) file does not exceed 10 minutes, follow step B. below and proceed to "Transcribing with CLOx."*

***Step B: Resampling***

Your .wav file should be sampled at 16,000 Hz. Resampling may be done in Praat by selecting the mono file and clicking Convert – Resample… and entering 16000 in the New Sampling Frequency (Hz) box. The precision does not need adjusting. Files sampled at less than 16,000 Hz may also be used, though transcription accuracy will likely decrease.

***Step C: Segmenting***

If you would like to manually segment your audio into extracts under 10 minutes in length or if you have multiple files that you would like included in the same transcript:

1. *Make sure that each extracted audiofile name ends with _”startTime”.wav, where “startTime” is the beginning time of the file in milliseconds. For example, if you are using 9.5-minute (570000 millisecond) extracts, the first .wav should be named soundName_0.wav, the second soundName_570000.wav, the third*

*soundName_11400000, and so on. (This is done to ensure that timestamps are represented accurately in the output .csv file.)*

2. *Create a unique local directory to hold the extracted .wav files that you want included in the same transcription.*
3. *Sort the extracted files by name, so the first file listed in the directory is soundName_0.wav, the second is soundName_570000.wav, etc.*

## TRANSCRIBING WITH CLOx

CLOx works with an open connection to a Microsoft Speech Recognition server. For best results, make sure that the screensaver and sleep (sometimes called energy saver) system settings of your computer are turned off while CLOx runs. You may, alternatively, set both to start only after an extended period of time (2 hours should be more than sufficient for even a lengthy CLOx session on an older computer). This ensures that you will not experience disconnection from the server due to periods of inactivity. You should also avoid navigating your internet browser away from the CLOx webpage while CLOx is running.

1. *Enter your subscription key in the "Subscription" field.*
2. *Select the language of your interview/audio file.*
3. *Enter a name for your output file.*
4. *Click "Select Files and Start."*
5. *A dialog box will appear. Navigate to the folder containing the preprocessed files. Select desired files using shift+click, ctrl+click, or cmd+click. Files should be sorted by name in ascending order. Press enter or click ok.*
6. *The results box will update to display CLOx's progress. It shows the transcription as it is being completed and indicates that text is*

*being added to the output file row by row. As a rule of thumb, allow up to one minute of transcription time for every one minute of speech in your audio file(s).*

7. *When complete, the comma-separated output file will be downloaded to the default folder for downloads on your computer, with the name you specified for output. This file may be opened in Microsoft Excel, or any standard text editor.*

8. *The output file contains 3 columns, labelled "Text," "Onset" and "Offset." an example is shown below:*

| | A | B | C |
|---|---|---|---|
| 1 | Text | Onset | Offset |
| 2 | Yeah no I mean I notice them sometimes | 3130.82 | 3133.84 |

The first column, "text," contains the speech recognizer output. The second and third, "onset" and "offset," respectively, contain the beginning- and ending-times of each "turn," determined by Microsoft's algorithm as a run of speech ending with a period of silence above an arbitrary threshold, or when Microsoft detects a change in speaker vocal quality. Each row indicates a "turn." CLOx appends all output to a single .csv file. So, regardless of whether a single audiofile or multiple audiofiles were selected, all output is concatenated into a single output file. This saves the user from having to concatenate multiple transcripts associated with a single interview session.

## WHAT DO I DO WITH MY OUTPUT?

While the Speech Service API provides highly accurate transcriptions, it is far from flawless. In addition to occasional transcription errors of words or clauses, CLOx is currently unable to reliably separate speakers on a recording and may have difficulty accurately transcribing overlapping or otherwise obscured speech. Therefore, it is

important to check and manually correct your transcriptions. We recommend importing your CLOx output to a software application such as ELAN, that allows auditing of your audio alongside your CLOx output. Here's how:

### Manual Correction in ELAN

CLOx transcripts are designed for easy importing in ELAN. Follow these steps:

A. In ELAN, instead of creating a new project, select File → Import → CSV / Tab-delimited Text File…
    1. Select the "Text" column as "Annotation"
    2. Select the "Onset" column as "Begin Time"
    3. Select the "Offset" column as "End Time"
    4. Specify first row of data: 2

B. Now add your audio file by selecting Edit → Linked Files…
    1. On the Linked Media Files tab (selected by default), click Add…, select the file containing the entire recording of your transcription, and click Apply.

You should see the waveform of the audio file appear above the transcription in the main ELAN window, and it should be properly aligned with your transcription.

## KNOWN ISSUES

*?? CLOx returned a blank .csv transcript*

There are three common reasons for this:

1. *There is a problem with your audio file. Ensure that it is preprocessed correctly, and then listen to the preprocessed audio files to ensure they contain the correct audio. Silent portions of your recording greater than two seconds each may also cause the transcription process to end prematurely, but the current version of the Praat preprocessing script will trim silences down to 1 second.*
2. *There is a problem with your subscription key or location and language settings. You may be using an old subscription key from the Bing Speech API; if so, you will need a new key for Microsoft Speech – see this guide for how. Check to make sure that you have not entered the subscription key incorrectly, and that you have specified the correct region and language settings. Your free trial may also have expired.*
3. *There are reports of free accounts not working for some users, so it is a good idea to upgrade to a pay-as-you-go account.*

*?? My connection terminated prematurely*

If you experience connection issues mid-transcription, you will need to refresh the page and start again from the beginning. You may wish to run transcriptions in segments if this error persists. For best results, any system settings (sometimes called energy saver) that automatically put the hard disk to sleep should be turned off while CLOx runs, and you should avoid navigating the browser away from the page while CLOx is running.

## CONTACT US

Questions? Issues? Contact Rob Squizzero at cloxhelp@uw.edu.

Thanks for using CLOx!

## SECURITY AND PRIVACY OF DATA

It is expected that many CLOx users are working with audio data that may contain subjects' personal or identifying information and may be subject to scrutiny by Institutional Review Boards (IRBs). To that end, we recognize that data security is of paramount importance and would like to explain that audio files and generated transcriptions never pass through or are intercepted by the CLOx server. CLOx, as a service, merely facilitates transfer between a user's computer and Microsoft's speech-to-text servers. Microsoft's policies prohibit their storage of any audio recordings or transcripts generated for any purpose.