Ghanem, R., Edalatishams, I., Huensch, A., Puga, K., & Staples, S. (2020). The effectiveness of computer programs in the transcription and analysis of spoken discourse: Towards a protocol for pronunciation corpora. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 97–114). Ames, IA: Iowa State University.

THE EFFECTIVENESS OF COMPUTER PROGRAMS IN THE TRANSCRIPTION AND ANALYSIS OF SPOKEN DISCOURSE: TOWARDS A PROTOCOL FOR PRONUNCIATION CORPORA

Romy Ghanem, University of Memphis
Idée Edalatishams, Iowa State University
Amanda Huensch, University of Pittsburgh
Karin Puga, Justus-Liebig University of Giessen
Shelley Staples, University of Arizona

The analysis of spoken language has always had to address two main issues: the tedious method of extracting variables and researchers' potential subjectivity when analyzing the data. The use of digital tools/software programs afford partial automatization of processes and the potential for increased objectivity. While these tools have improved analysis processes, some programs are more suitable for certain aspects of analysis than others. This paper provides an overview of five digital tools that have been recently employed in the analysis of spoken data: CLAN, ELAN, MFA, Phon, and Praat. The aim is to offer a brief but practical introduction to the usefulness of each program in relation to: transcribing audio files and extracting segmental (e.g., vowel length), prosodic (e.g., pitch range), and fluency (e.g., pause length) features. We hope that our protocols for transcription and analysis can also be used to further the ongoing goal of developing pronunciation corpora for use by the PSLLT community. The paper provides links to open source materials and information about each program through our website. We invite others to contribute to the website with their own handouts and resources.

INTRODUCTION

The analysis of spoken data frequently presents two main obstacles to be addressed: the tedious nature of analyzing data, and potential subjectivity. In the last few decades, developments in computer software (such as CLAN, ELAN, MFA, and Praat, among others) have allowed for the automatic or semi-automatic extraction of certain phonological features, which in turn could possibly lead to increased objectivity. With the recent boom in digital tool development, a pronunciation researcher may find it quite difficult to choose the most appropriate software for their data and various types of analyses.

Alongside these developments, there has been a recent call for systematicity in the creation of spoken corpora (Allwood et al., 2002; Cermak, 2009). One argument supporting systematicity stems from the need to compare two similar spoken corpora. If different annotations or coding schemes are used, this is not possible (Allwood et al., 2002). The Pronunciation in Second Language Learning and Teaching community has also expressed interest in creating a multilingual corpus that can be open to all pronunciation researchers. Specific tasks related to building, transcribing, and annotating the corpus have been brought up in recent workshops, particularly at the 10th PSLLT conference by Shelley Staples and Amanda Huensch.

Choosing or compiling a corpus is an initial step that precedes the steps described in this paper. After our initial discussions with PSLLT members in 2018, we do not think there is one corpus that will meet all of the needs of members. As Huensch and Staples (2018) and others have discussed, one of the main limitations of existing spoken corpora for pronunciation is their lack of sound files. We therefore encourage open data sharing and collaboration among PSLLT colleagues. Alongside that open data sharing, we hope that this paper provides guidance on protocols for data segmentation, transcription, and analysis that can be used across corpora in the PSLLT community.

For this paper, we were able to use the Corpus of Collaborative Oral Tasks (CCOT) (Crawford, in preparation) which has been made available to a number of second language researchers including those in the field of corpus, pronunciation, and assessment (see DeGruyter: *Multiple perspectives on learner interaction: The corpus of collaborative oral tasks* edited by William J. Crawford, in preparation). Because the corpus compilers are able to make the sound files available to researchers, we would like to suggest this corpus as a possible data source for PSLLT researchers interested in the intersections between pronunciation features and task and/or language proficiency in dialogic interactions among L2 learners of English. The tasks are fully described and each interaction has received a language rating. L1 background is included in the speaker metadata, making this another possible variable to explore. Those interested in using the CCOT corpus are encouraged to contact the creator, William Crawford, at Northern Arizona University.

PURPOSE

The purpose of this paper is to briefly introduce five computer software tools that pronunciation researchers could use in the transcription, annotation, coding, and analysis of spoken data. The speech file used to demonstrate the functions in each program is a 30-second clip from a dialogic interaction between two second language speakers of English, part of the CCOT (Crawford, in preparation) corpus described above. In the interaction we investigated, the two students engaged in an informative task where they were asked to discuss graphs and data related to crime and economy.

The five digital tools introduced in this paper are: CLAN, ELAN, MFA, Phon, and Praat. The strengths and weaknesses of these programs are highlighted in relation to the most common steps researchers perform when analyzing speech files: data preparation, data coding (including segmental and suprasegmental features), and data analysis (see Figure 1 for a schematic representation). The interoperability among the five tools as well as their usability will also be addressed. In the process of analyzing these software, detailed handouts for their use were created. These are available on our presentation website: https://sites.google.com/view/psllt2019/home. More details on the handouts and website are discussed at the end of the paper.

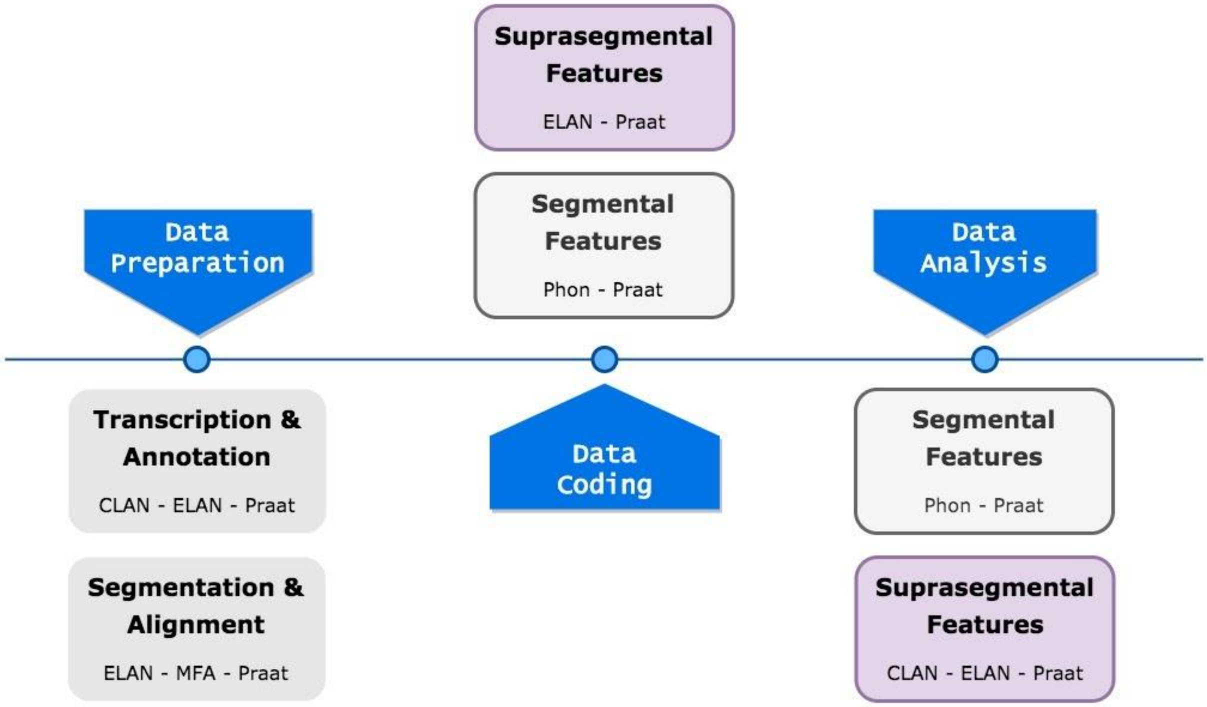

Figure 1. Possible digital tools at various stages of spoken data research

In the following sections, we define transcription as the process by which speech is represented through text or phones, whereas annotation is the addition of any segmental (e.g., VOT) or suprasegmental (e.g., pitch values) information to the text. In order to conduct fluency and pronunciation analysis, segmentation is most often employed, whereby spoken discourse is separated from pauses. This could be done manually or automatically as will be explained below. Alignment is defined as the technique through which a program temporally matches sound with transcriptions (whether phonetic or orthographic). We refer to data coding as the procedure of extracting segmental (at the consonant and vowel level) and suprasegmental (above the segment) features. Data analysis, on the other hand, is the statistical representation of the data (e.g., means, maximal/minimal values, etc.) as produced by the digital tools discussed.

DATA PREPARATION

Manual Segmentation

ELAN: EUDICO Linguistic Annotator (Max Planck Institute for Psycholinguistics, 2018) is a computer software that allows for the manual transcription and annotation of audio and video files. ELAN is highly compatible with other programs used for pronunciation analyses. TextGrid files can be imported from Praat, and ELAN files (.eaf) can be exported to be further analyzed in Praat. Praat (Boersma &Weenink, 2019) is a freely available speech analysis software for phonetic and phonological analyses. It can be used to view and annotate audio files (mainly WAV files) and TextGrids. A TextGrid file is a text used to identify certain orthographic or pronunciation features in a sound file. Multiple layers of transcription and annotation can be included in a TextGrid through tiers (see below for more explanation). CHAT files, which consist of a transcription aligned with an audio file, can also be imported from CLAN and analyzed in ELAN.

The first step to take in ELAN is to create Tiers. A tier could be thought of as a collection of annotations (runs and pauses) that share some characteristics. When analyzing spoken (in this case, dialogic) speech, a tier often represents a speaker's productions (either silent or vocalized). For instance, one can create a tier for Speaker A and also another one for Speaker A Pausing. One way to segment speech in ELAN is to click on the segmentation mode tab and manually divide various annotations based on the speaker's speech or pausing. The lines around each annotation turn green when the mouse pointer is hovering over it, which lets a coder modify the length of that annotation. Figure 1 demonstrates four tiers that represent the speech and pauses produced by the two speakers. The tier that is being modified appears at the top of the list in red, which allows for the expansion or retraction of each annotation. One cannot add or modify text in segmentation mode, but creating a tier, modifying an annotation, and aligning tiers are possible in this mode. While this segmentation process is relatively easy, other computer software (such as Praat) allow for the automatic segmentation (with some revisions necessary) of speaking time vs. pauses. This could save the researcher quite a bit of time compared to doing so in ELAN.

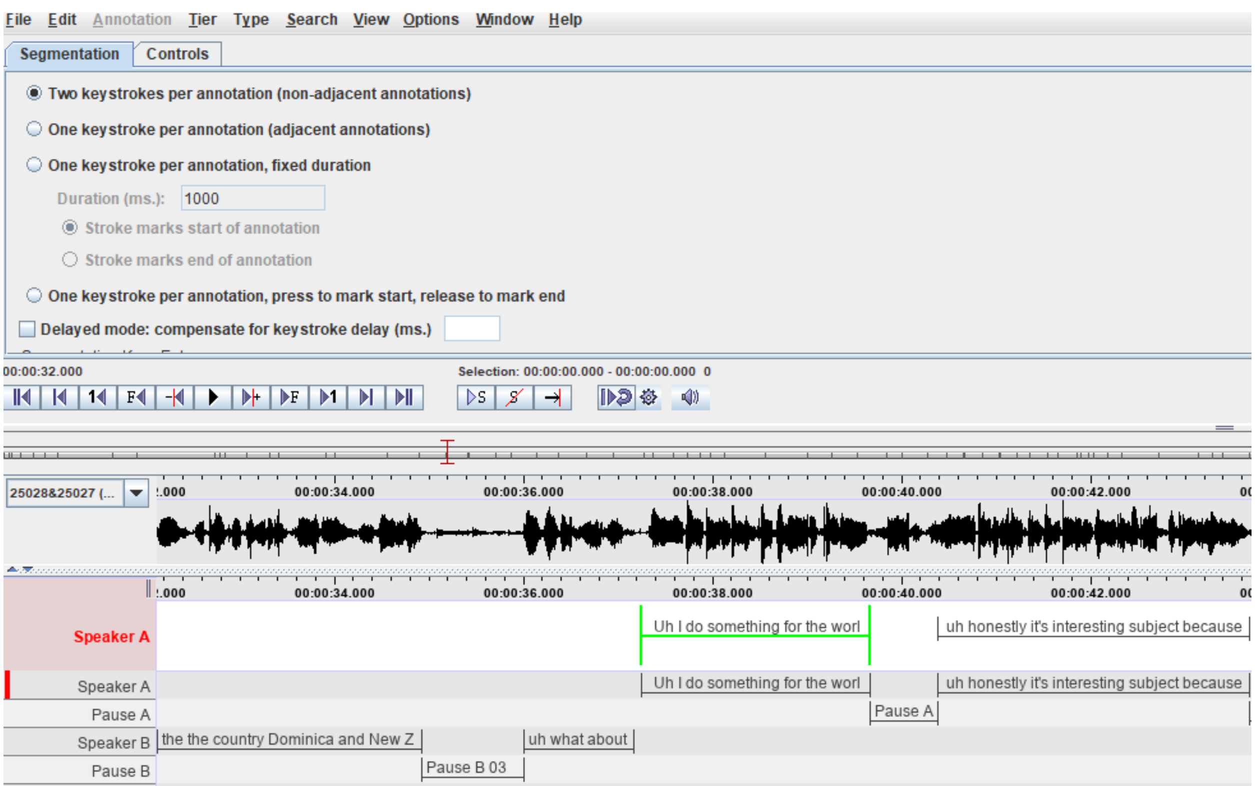

Figure 2. Segmentation mode in ELAN

Transcription

ELAN

For already transcribed files such as the one we used, one can simply copy and paste text into the annotations that were created in the previous step. ELAN includes a feature which gives access to the speaker's entire speech (and pauses if one chooses). To do so, one can access the Transcription mode through the options menu and access the various tiers being annotated. Every annotation is numbered as shown in Figure 2, and one can play the individual sections to ensure accuracy. Transcription mode can also be used to transcribe text as the coder is listening to each annotation. The loop mode allows for repeating each section. The rate of speech can also be adjusted, which facilitates the transcription process. Transcriptions can be exported from ELAN to NotePad as .txt files.

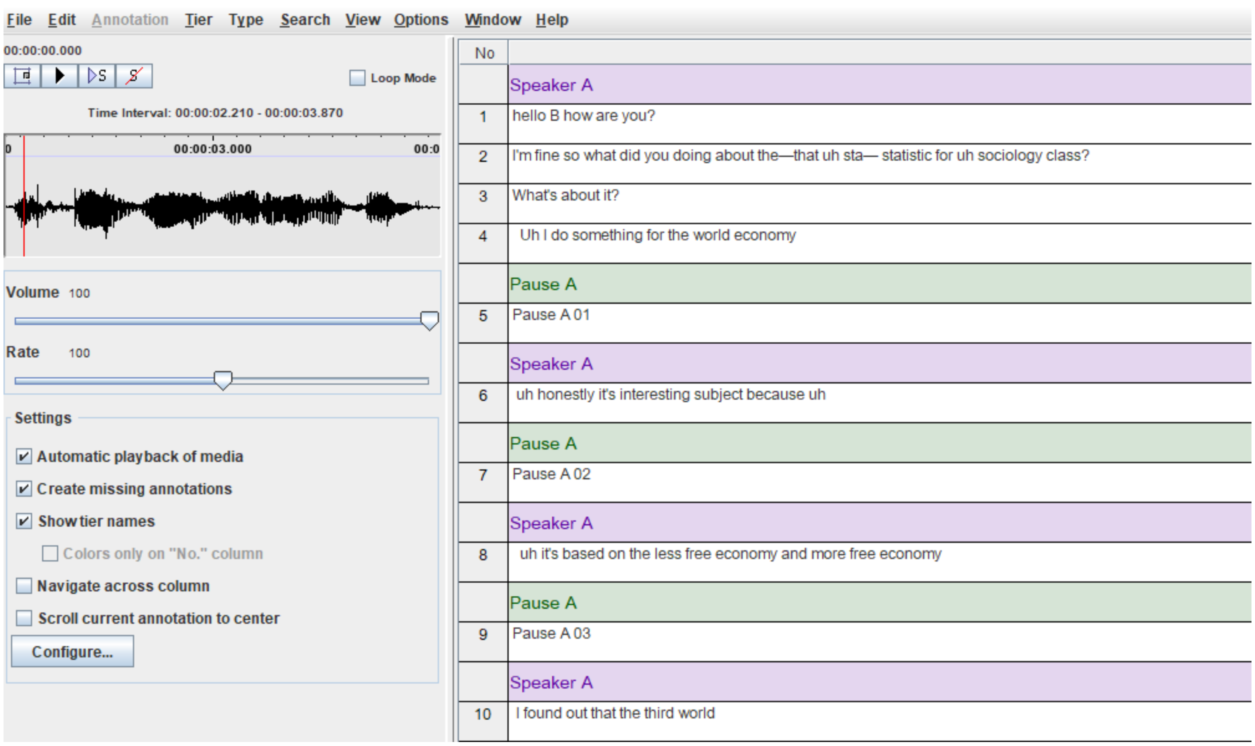

Figure 3. Transcription mode in ELAN

Another way to transcribe speech in ELAN is to do so in Annotation Mode (see Figure 4). Here, users can first create a new annotation (or use a segmentation program such as Praat to first create the annotation boundaries) and then type directly into the annotation area.

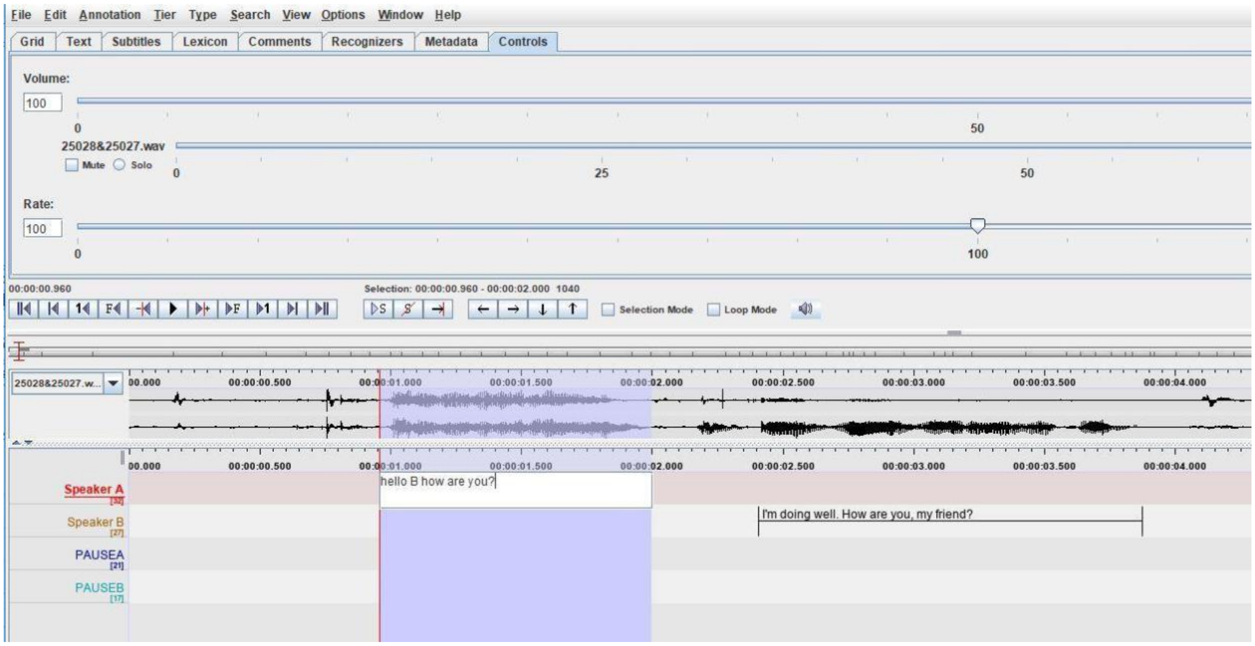

Figure 4. Annotation mode in ELAN

CLAN

CLAN (Computerized Language Analysis; MacWhinney, 2000) is a computer software that allows for transcription and analysis of corpora and was developed as part of the TalkBank system. It is designed to be used with data which have been transcribed following universal CHAT transcription conventions. CLAN has two main components: (a) a text editor which facilitates the transcription of audio and video media, and (b) a suite of data analysis programs that can perform a variety of corpus analyses on transcribed and coded data. One benefit of CLAN is that it has a wealth of documentation and a strong history of support. The TalkBank web site contains manuals for CHAT transcription and CLAN use that are thorough. Another advantage of CLAN is that it benefits from high interoperability with other commonly used software programs for speech analysis (e.g., ELAN, Phon, Praat). It facilitates easy conversion between file types that allows for each program's strengths to be capitalized upon.

Figure 5 illustrates the CLAN editor window including metadata headers (marked with @), the transcript, and the sound file waveform. The header in Figure 5 is relatively basic but still includes information about the languages in the transcript, the speakers, the date of data collection, the task topic, etc., which facilitates corpus linguistics. The CLAN commands window, within which many functions and analyses of the CLAN programs are executed, is present to the right of the transcript. Transcripts can be time-linked to audio and video data (indicated in Figure 5 by the bullet points at the end of each utterance line). CLAN supports a variety of audio (WAV, AIFF, AIFC, and MP3) and video (e.g., MOV, MPEG4) formats. More information about supported formats and transcription techniques can be found in MacWhinney and Wagner (2010).

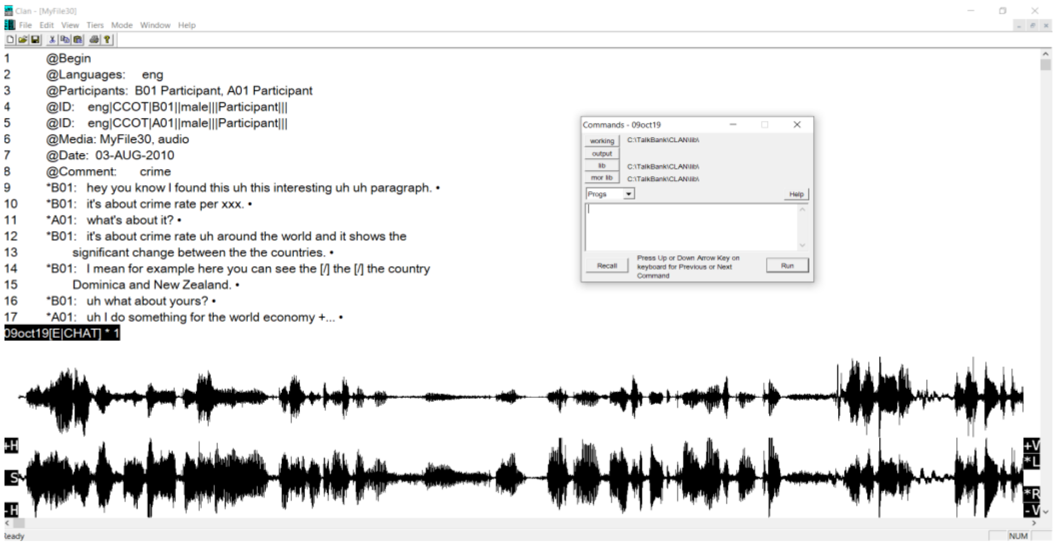

Figure 5. Example transcript in CLAN editor

Automatic Segmentation

Praat

Praat is well-documented with online step-by-step guides and video tutorials. While Praat is user-friendly, it involves a time-consuming manual data coding and data extraction process. For more complex data segmentation and extraction tasks, Praat scripts or plug-ins are needed. However, one feature that is part of Praat's standard settings is the auto-segmentation of sounding and silent intervals within a given speech file (see Figure 6 for an example).

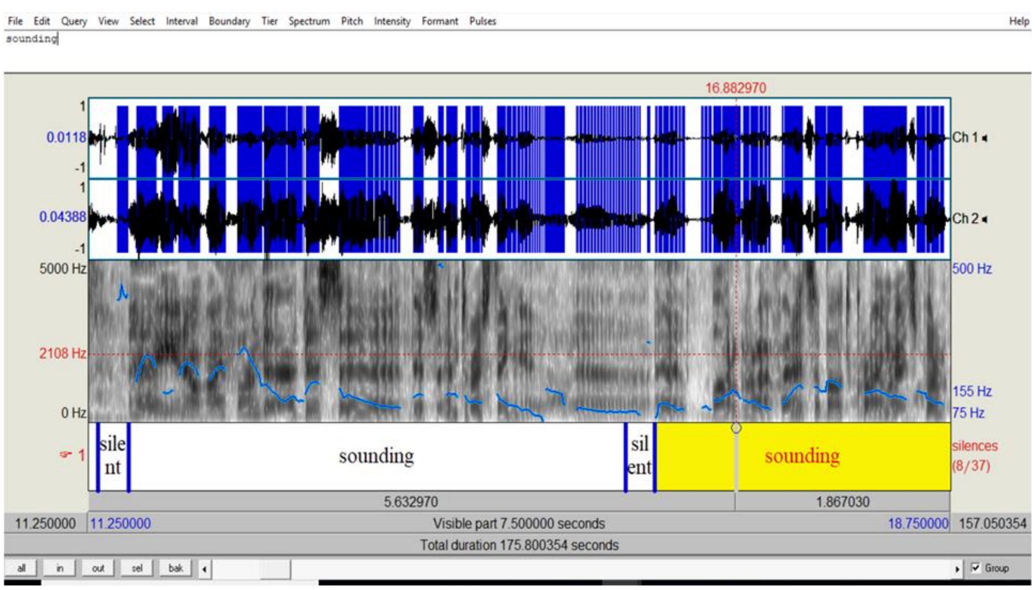

Figure 6. Praat screenshot of sounding and silent intervals - automatic-segmentation

Once the boundaries are set, one can click on the grey bar below each interval to verify auditorily whether boundaries have been set correctly. Intervals can be corrected by dragging the blue vertical lines to their respective place. The “sounding” intervals can then be replaced by available transcripts by copying and pasting the text in the respective intervals. An alternative to this standard Praat feature is using an EasyAlign plug-in available for Windows (Goldman, 2011) that automatically aligns a given transcript with an audio file in Praat. For such a procedure only an audio file and transcript are needed. The utterances will be segmented according to each line of the transcript. The researcher should thus prepare beforehand a text file with each line representing an utterance to be segmented into an interval. A sample annotation with EasyAlign can be seen in Figure 7.

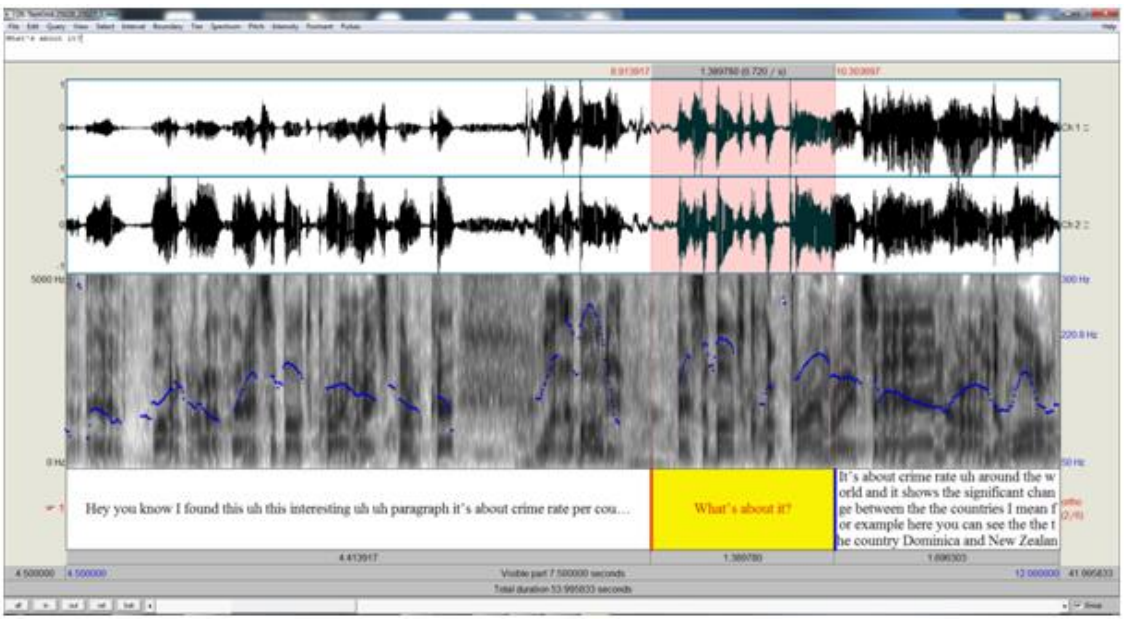

Figure 7. Praat screenshot with automatic utterance segmentation with EasyAlign

Automatic Alignment: Montreal Forced Aligner

MFA (Montreal Forced Aligner; McAuliffe et al., 2017) is a freely available software for aligning orthographic transcriptions with audio files and creating a time-aligned version of the transcripts. It makes use of a pronunciation dictionary to look up phones for words. The software does not have a user interface but comes in a package for installation and needs to be run in the terminal window on Mac operating systems or the command window on Windows computers. The software is supported with thorough online documentation from the producers. It is recommended to first align an example dataset (available on the MFA website) to make sure one is running the scripts correctly before running the aligner on your own data.

MFA requires some data preparation before it performs the alignment. The transcript files and audio files need to be saved with the same name and stored in the same directory. The transcript files can be in either TextGrid, TXT, or LAB formats, but the only accepted format for the audio is WAV. For best results, the audio files need to be shorter than one minute long. An example of our prepared short audio file and its transcription in TextGrid format can be seen in Figure 8. Additionally, an output directory needs to be created for the resulting TextGrid files generated by MFA.

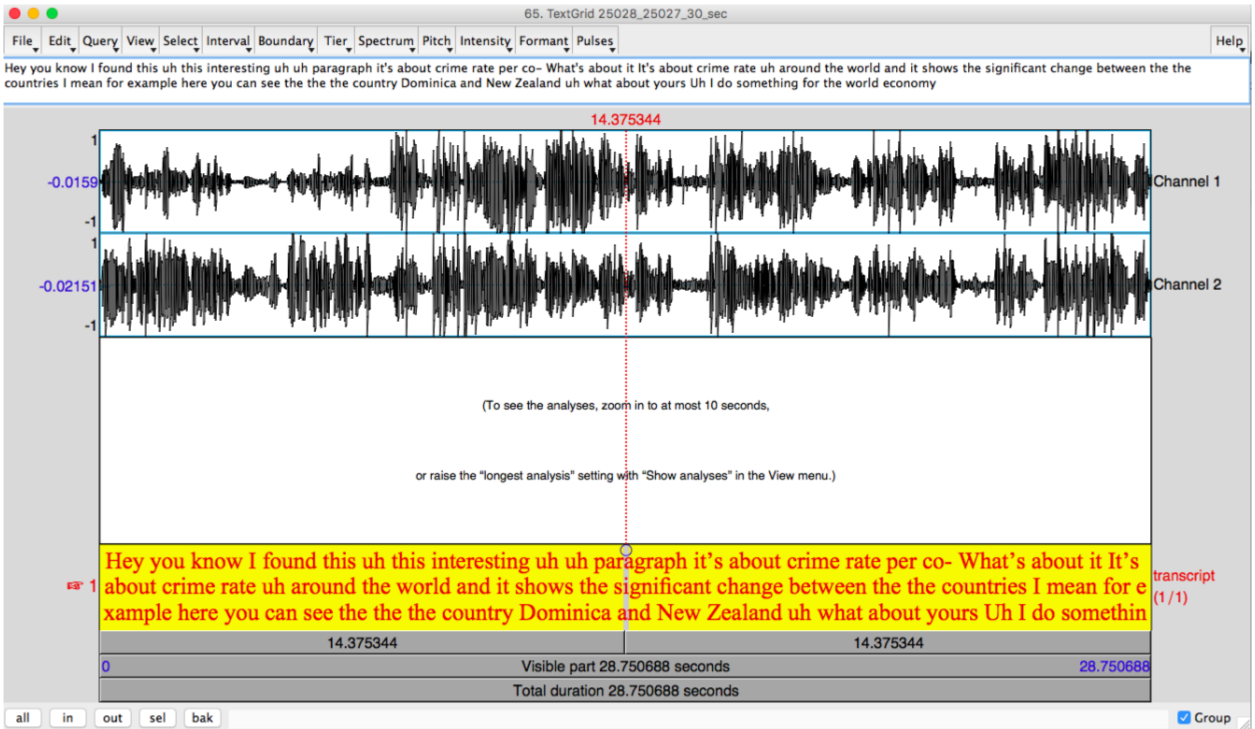

Figure 8. Sound and transcription example before alignment

The user also needs to install the MFA package and download a pre-trained acoustic model before running the MFA program on the data. Figure 9 shows an example of MFA run in the Terminal window on a Mac computer.

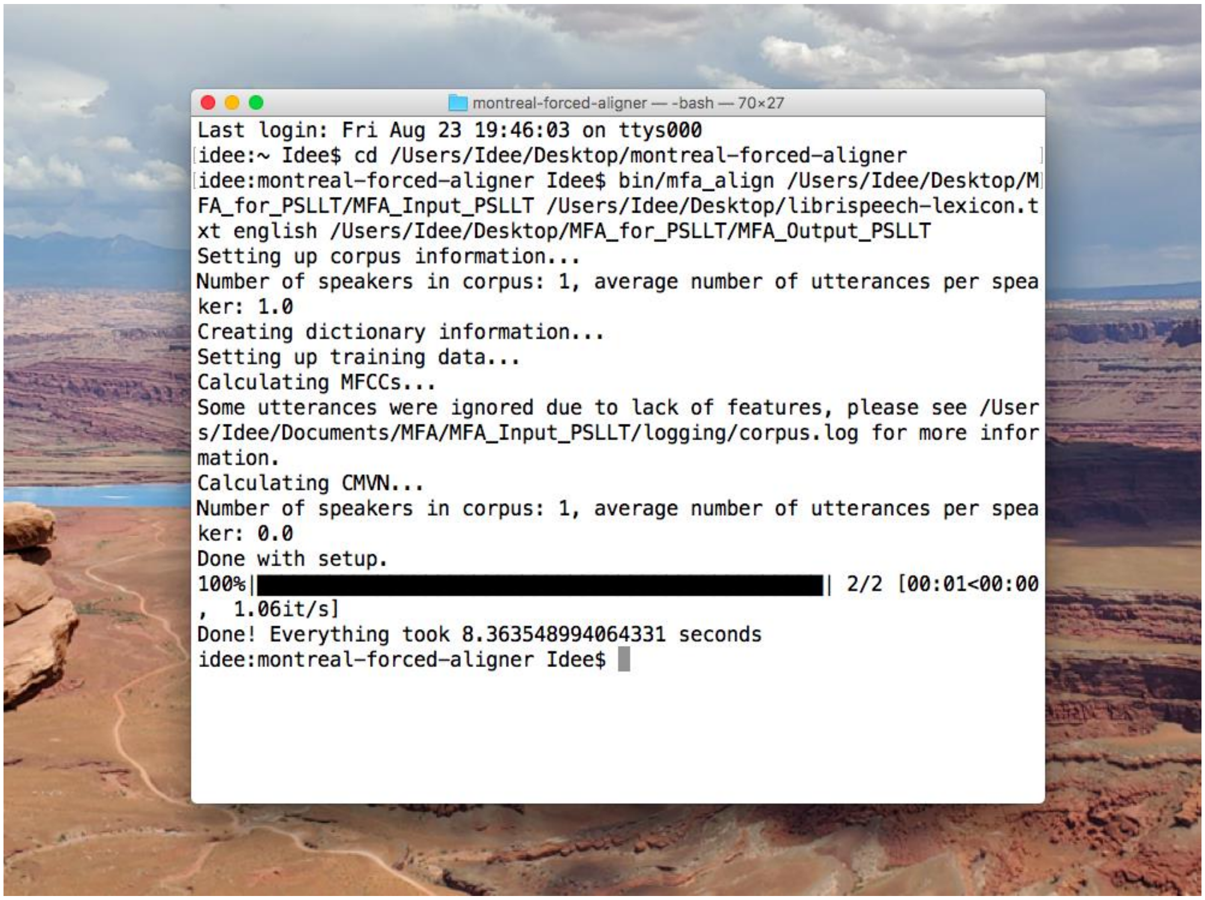

Figure 9. MFA run in the terminal window on a mac computer

Running the program results in generating new TextGrid files stored in the output directory. Figure 10 displays a Praat window with transcripts aligned with the audio at the phone and word levels in two different Praat tiers. These files will have the same names as the original files and correspond to the original short audio files. Users can combine the short audio files and concatenate the transcription files in the end in Praat.

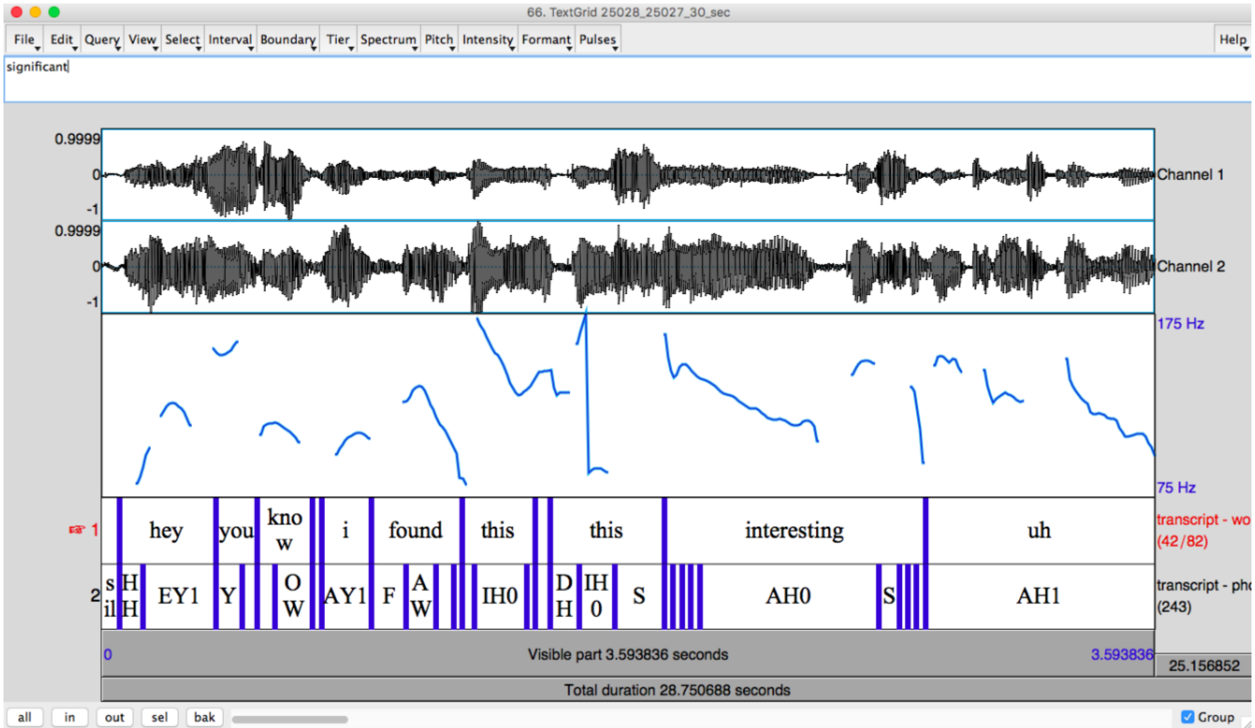

Figure 10. Sound and transcription example after alignment

MFA performs well with different types of data, such as read or spontaneous speech. Its trainability, interoperability with Praat, and compatibility with Prosodylab-Aligner (Gorman et al., 2011) and other formats are some of the strengths of this aligner. MFA works with many languages, allows for aligning speech by multiple speakers in different tiers, and accounts for silence, noise, or unknown words. However, the accuracy of the alignments heavily depends on factors including audio quality, transcription accuracy, and accentedness. For example, for the audio file used in our analysis, although MFA aligned the speech with background noise, it had some difficulty with overlapping conversations. MFA can produce a report of issues that need to be fixed before the alignment but may skip aligning some of the files without reporting a reason if some issues are not identified and/or fixed.

DATA CODING

Segmental Features

Phon

Phon (Hedlund & Rose, 2019) is a software program for building and analyzing phonological data corpora (see Figure 11). It was built to interact with CLAN and also has interoperability features with Praat. Advantages of using Phon include being able to coordinate multi-blind IPA transcription, automatically convert orthographic transcription to (idealized) IPA transcription, and conduct automatic syllabification. Similar to CLAN, Phon can be used with a variety of audio and video file formats (e.g., MP3, WAV, MOV). As we attempted to use Phon with the CCOT file, we encountered some difficulties with the functionality of the program, and given that Phon is a relatively new program, less documentation for it exists online. Nevertheless, the programs' creators were generous with their time when support was requested on the program's Google Group (Phon). We see future potential in this program for pronunciation researchers, but also the need for more documentation and training.

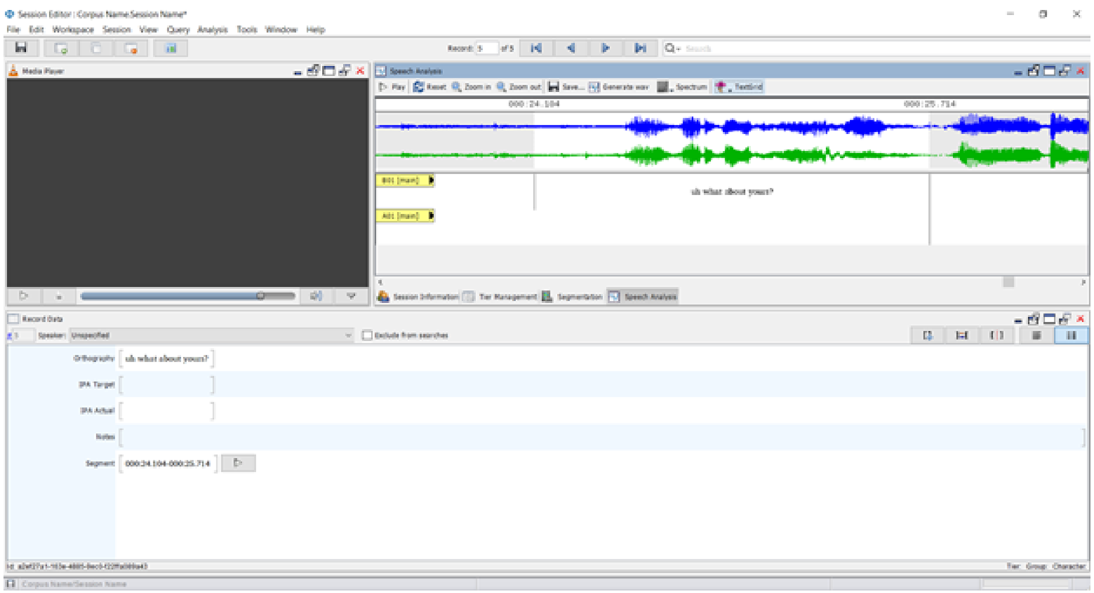

Figure 11. Screenshot of Phon

Praat

In Praat one can create new tiers for segmental features: pauses, syllables, words, etc., and an inter-rater agreement plug-in provides the user with inter-rater agreement statistics for syllables, for instance. The number and type of tiers should be decided upon by the researcher from the outset, because deleting or adding tiers at a later stage can require much time, especially if many audio files are to be annotated. Figure 12 shows an example of a manual annotation of segmental features, including three interval tiers.

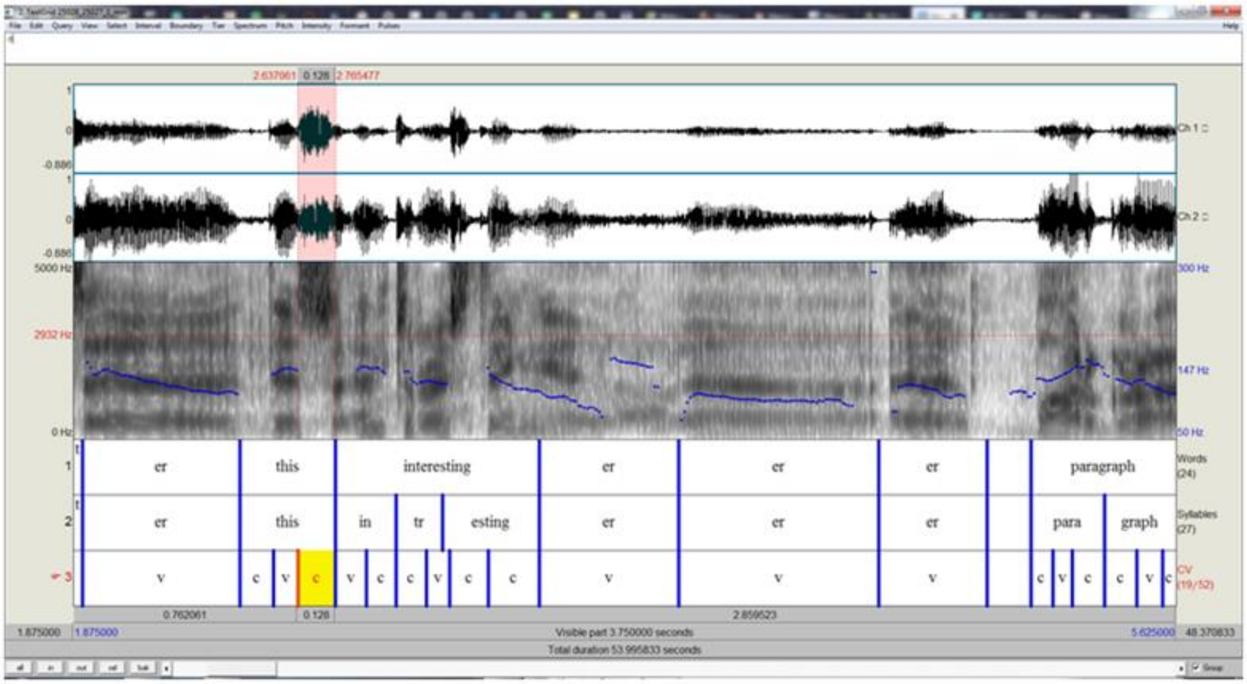

Figure 12. Praat screenshot of a manual segmental annotation including a word tier (1), syllable tier (2), and a consonant/vowel tier (3).

Visual inspection can be undertaken by investigating the spectrogram. The degree of grey-shading can help determine the specific consonants and vowels. The sound "s" for instance, as selected in the tier in Figure 12 (highlighted in yellow on the tier and pink in the waveform), is often a dark grey area in the upper part in the spectrogram.

Suprasegmental Features

Fluency features: ELAN

ELAN is a great tool to use for coding fluency features, particularly pauses and runs. As shown in Figure 2 above, pauses (silent or filled) can be coded on separate tiers. If the data has been pre-segmented by another program such as Praat into silent pauses and runs, the program can easily be used to automatically or manually copy the pause segments to a new tier and analyze them separately.

Prosodic features: Praat

Praat is an excellent tool for prosodic data coding. Prosodic features such as pitch, intensity, formants, and pulses can be analyzed. Praat enables the researcher to visually inspect these prosodic features and to annotate the speech signal on multiple point- and interval tiers, using the prosodic annotation system of their choice (e.g., Tone and Break Indices [ToBI] or Brazil's discourse model; Brazil, 1997; Silverman et al., 1992). An example of a manual ToBI annotation is illustrated in Figure 13.

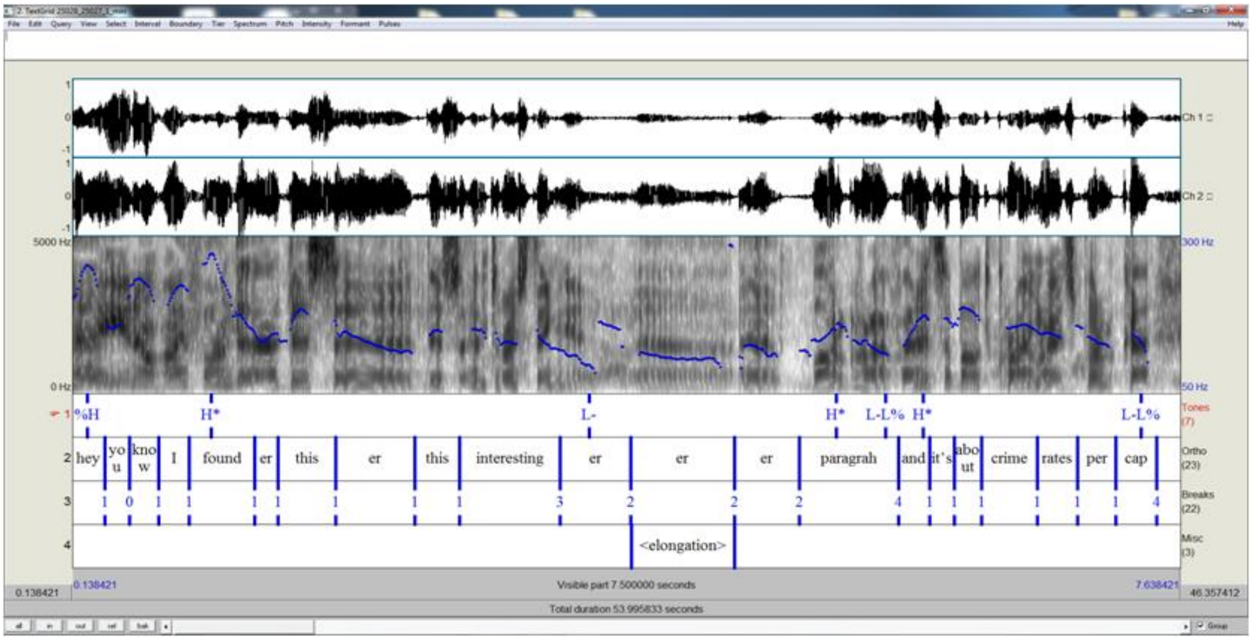

Figure 13. Praat screenshot of sample ToBI notation system

Figure 13 includes the four tiers a ToBI annotation should minimally consist of, i.e., a tone tier (point tier; 1), an orthographic tier (interval tier; 2), a break index tier (point tier; 3), and a miscellaneous tier (interval tier; 4). This is one of the advantages of Praat: It enables researchers to choose between different prosodic notation systems (e.g., Brazil's discourse model, Halliday's nuclear tones, ToBI) and adjust them for their purposes.

DATA ANALYSIS

Segmental Features

CLAN

Data that have been transcribed and annotated in CLAN can take advantage of several programs for automatic analysis of fluency features. To run these programs, it is first necessary to conduct a MOR (morphological) analysis, which provides part of speech tagging for words in the transcript. The MOR analysis automatically provides syntactic information about each of the words in the transcript. Once MOR has been run, other programs use this information to automatically calculate fluency features. One program is called MLU (mean length of utterance), and it can be run to extract information about mean length of utterance (i.e., the average number of morphemes per utterance). CLAN also includes multiple options for customizing the MLU analysis. Another useful program is called FLUCALC (Fluency Calculation). This program can be used to automatically output over 30 measures of speed, breakdown, and repair fluency (Tavakoli & Skehan, 2005).

ELAN

ELAN allows for the automatic calculation of some fluency features used in pronunciation studies including average pause length, number of pauses, average run length, and speaking time (per speaker). One can choose to extract these measures by speaker or by tier. Doing so by tier would help the researcher to collect information about the speakers' speech time and their pauses (number and length). To do so, one can simply access the Statistics tab and the information will be displayed in a dialog box that can be later transferred to CSV format. Figure 14 displays the statistics for both speakers in our speech file. ELAN's interoperability with other programs allows for easy transfer of data to calculate other suprasegmental features.

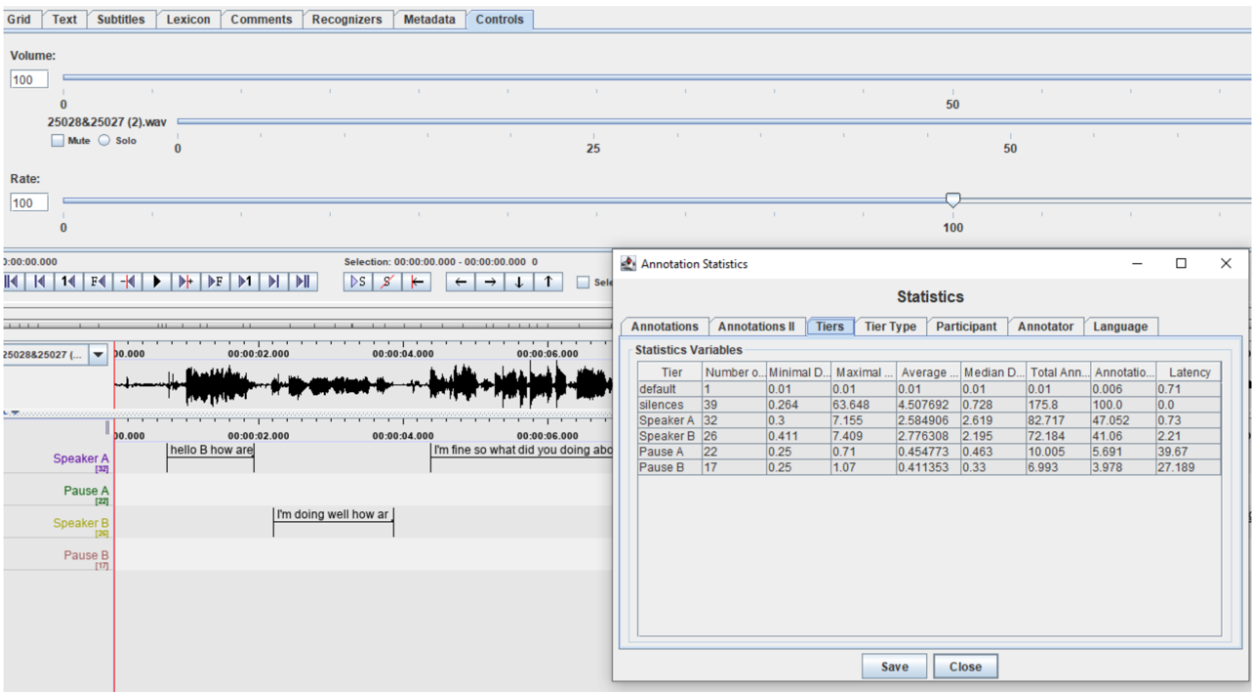

Figure 14. Suprasegmental feature values in ELAN

Phon

Once data have been transcribed in IPA, Phon is designed to facilitate phonological analyses. For example, Phon queries allow for search parameters that differentiate between features (e.g., nasal, coronal) and syllable and word positions. Rose and MacWhinney (2014) described an analysis of the acquisition of velar stops by child L1 learners that was able to take into consideration both the location of the velar (i.e., initial, medial, final) and syllable stress (i.e., primary, secondary, unstressed) (pp. 19-20). These built-in functions of Phon allow for more focused queries.

Praat

Within Praat it is possible to get automatic information on each interval, i.e., the space between two boundaries (vertical blue lines) on a Praat tier or between each mark on a point tier (see Figure 13). For instance, by selecting an interval one can get information on the length of the interval as well as its Hertz level. However, selecting each interval individually can be very time-consuming. Freely available Praat scripts for frequently used functions, such as counting the frequency of all intervals on a specified tier, can be found online and implemented by users (e.g., http://phonetics.linguistics.ucla.edu/facilities/acoustic/praat.html). Another option would be to open the TextGrid file in ELAN; however, only interval tiers will be transferred from TextGrid annotations. Point tiers, on the other hand, will not be accessible in ELAN.

IMPLICATIONS

The PSLLT community has shown interest in creating a multilingual corpus that can be accessed by researchers with different theoretical frameworks and research designs. Being aware of various digital tools and ways of using them could be the first step towards developing a protocol for annotating and examining existing pronunciation corpora. The purpose of this paper was to briefly introduce five digital tools used for pronunciation research: CLAN, ELAN, MFA, Phon and Praat. More detailed handouts have been uploaded to our presentation website: https://sites.google.com/view/psllt2019/home. The individual protocol handouts for each program detail step-by-step procedures starting from program installation to descriptions of the various analyses possible with each program. A spreadsheet was also created for a quick visual summary of the skills and features that are available for each digital tool and the usability of each program as it relates to accessibility, ease of use, and compatibility with other programs. Our website is open for other pronunciation researchers and experts to upload their own protocol handouts of different digital tools that can be equally useful for analyzing spoken data. We encourage other pronunciation experts to contribute similar protocols for other digital tools they have used. To contribute handouts for other programs, visit our presentation website for more information. Although most programs have ample online resources, specific materials designed by individuals who deal primarily with speech files would offer a new and more relevant perspective on the usefulness of the tool. Our website also includes other resources that could be beneficial for researchers who deal with speech data, including a list of available spoken corpora as discussed in the 2018 PSLLT colloquium led by Amanda Huensch and Shelley Staples. The listed corpora can be used by various researchers with different research questions and goals.

To support the PSLLT community, researchers who are compiling their own corpora are invited to review the IRB documents provided by Huensch and Staples (2018) (also available on our website) as guidance in collecting data that can be used by the PSLLT community. Other steps that need to be taken after those described in this paper include accuracy checking of the variables identified in the data analysis. Intercoder/interannotator agreement and/or human/computer annotation agreement needs to be calculated to ensure the precision (ability to precisely identify a variable) and recall (ability to identify that the variable exists in the data). Procedures and protocols for checking accuracy are well documented for lexico-grammatical features (see, e.g., Biber & Gray, 2013), but procedures and protocols for conducting these kinds of checks would be a helpful next step for pronunciation data, particularly as more and more researchers use automated (Praat) scripts.

ABOUT THE AUTHORS

Romy Ghanem is an Assistant Professor of Applied Linguistics in the English department at the University of Memphis. Her studies explore topics in speech perception and production, applied phonology, and second language acquisition. Her research focuses mainly on second language speech, interphonology, and different varieties of a language. Her projects have examined issues related to nonnative speaker accentedness, the acquisition of second language phonology, ESL pronunciation instruction, linguistic stereotyping, and phonological convergence.

Idée Edalatishams is a PhD student in Applied Linguistics and Technology at Iowa State University. She uses corpus methodology to study native and non-native English speakers' pronunciation in natural discourse contexts. Her current research is focused on developing the prosodically annotated Corpus of Teaching Assistant Classroom Speech (CoTACS) and analyzing International and American TAs' use of intonation in their classroom discourse.

Amanda Huensch is Assistant Professor in the Department of Linguistics at the University of Pittsburgh. Her research uses corpus linguistic research methods to examine second language speech development in and outside of the classroom, including the relationship between perception and production, pronunciation attitudes of classroom foreign language learners, and fluency development during study abroad.

Karin Puga is a Ph.D. student and research assistant at the Justus Liebig University in Gießen, Germany. Her research makes use of corpus linguistic methods and tools to investigate L2 prosodic attainment by advanced learners of English. Her current research involves a corpus-based contrastive (interlanguage) analysis of the L2 prosody of Czech, German, and Spanish learners of English as compared to two native varieties.

Shelley Staples is Associate Professor of English at the University of Arizona. Her research focuses on corpus-based analysis of writing and speech. Her publications on corpus and pronunciation can be found in *English for Specific Purposes Journal* and in the volume *Spoken Corpora and Linguistic Studies*, published by John Benjamins.

REFERENCES

Allwood, J., Henrichsen, P. J., Grönqvist, L., Ahlsén, E., & Gunnarsson., M. (2002). Transliteration between spoken language corpora: Moving between Danish BySoc and Swedish GSLC. *Gothenburg Papers in Theoretical Linguistics, 86.*

Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL ibt® test: a lexico-grammatical analysis. *ETS Research Report Series*, *2013*(1), i-128.

Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.01, retrieved 14 August 2019 from http://www.praat.org/

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge University Press.

Cermak, F. (2009). Spoken corpora design. *International Journal of Corpus Linguistics 14*(1), 113-123.

Crawford, W. (under contract). Multiple perspectives on learner interaction: The corpus of collaborative oral tasks. New York: DeGruyter.

ELAN (Version 5.4) [Computer software]. (2018). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Goldman, J.-Ph. (2011). *EasyAlign: an automatic phonetic alignment tool under Praat.* Proceedings of InterSpeech, Firenze, Italy.

Gorman, K., Howell, J. & Wagner, M. (2011). Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech. *Canadian Acoustics*, *39*(3), 192–193.

Hedlund, G., & Rose, Y. (2019). Phon 3.0 [Computer Software]. Retrieved from https://phon.ca.

Huensch, A., & Staples, S. (2018). "Towards a protocol for a multilingual corpus for pronunciation researchers." PSLLT Workshop, Ames, IA.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner (Version 0.9.0). Retrieved from http://montrealcorpustools.github.io/Montreal-Forced-Aligner/

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion*, *11*, 154-173.

Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and Software-Assisted Methods for the Study of Phonology and Phonological Development. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (Eds.), *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press. 308-401.

Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, Janet Pierrehumbert, & Julia Hirschberg (1992): "ToBI: A standard scheme for labeling prosody", *Proceedings of the 2nd International Conference on Spoken Language Processing*, Banff, 867-879.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-276). Philadelphia/Amsterdam: John Benjamins.