1. INTRODUCTION

Audio speech corpora, such as the Audio British National Corpus (Coleman et al., 2012), the Buckeye Corpus (Pitt et al., 2005), and the Santa Barbara Corpus of Spoken American English (Du Bois et al., 2000), are common sources for phonetic study; however, few papers describe the steps necessary to build and analyze those corpora. Here we detail methods for transcription and forced alignment of a large-scale historical audio corpus, the Digital Archive of Southern Speech (DASS) (Kretzschmar et al., 2013). This collection of 64 sociolinguistic interviews (372 hours) is sampled from the extensive Linguistic Atlas of the Gulf States (LAGS) (Pederson et al., 1986), and features semi-spontaneous speech data from eight U.S. southern states, recorded 1968–1983. A limited number of target lexical items were originally impressionistically transcribed by LAGS fieldworkers and trained phonetic scribes; however, in the absence of full orthographic transcriptions, large-scale phonetic analysis has not heretofore been possible.

We begin by describing the DASS project and laying out our motivations for exploring this audio corpus. Next we describe our methods for large-scale transcription including preparing the audio, choosing appropriate transcription software, composing in-house transcription protocols as well as an in-house dictionary of non-standard words, training undergraduate transcribers, and taking steps to ensure consistency across transcriptions. Finally, we evaluate four methods of formant extraction of the resulting time-aligned orthographic transcriptions: DARLA (Reddy and Stanford, 2015), FAVE-Extract (Rosenfelder et al., 2014), an in-house modification of FAVE-Extract, and an in-house Praat (Boersma and Weenink, 2015) script. By transcribing and aligning the DASS corpus, we hope to bring new attention to this valuable archive of Southern U.S. speech. Furthermore, these methods may serve as a model to other projects that work with audio corpora.

2. THE DIGITAL ARCHIVE OF SOUTHERN SPEECH (DASS)

A. DASS AND THE LINGUISTIC ATLAS OF THE GULF STATES (LAGS)

The Linguistic Atlas of the Gulf States (LAGS) consists of sociolinguistic interviews conducted between 1968 and 1983 across the southeastern states of Texas, Louisiana, Arkansas, Mississippi, Tennessee, Alabama, Georgia, and Florida (Pederson et al., 1986). The original reel-to-reel and cassette recordings of LAGS' 1121 speakers have been digitized and maintained by the Linguistic Atlas Project (LAP) at the University of Georgia (Kretzschmar, 2011). Following in the tradition of previous Atlas projects such as the Linguistic Atlas of New England (LANE) (Kurath et al., 1939) and the Linguistic Atlas of the Middle and South Atlantic States (LAMSAS) (Kretzschmar, 1994), LAGS researchers sought to interview members of rural and urban communities who varied in both age and social background. Informants of European American (EA) descent were categorized based on the following social "types": Type 1, "Folk" speakers who were generally less educated with few community ties and social engagements; Type 2, "Common" speakers who typically had at least some high school education and social connections; and Type 3, "Cultivated" speakers, who usually had some college or equivalent education, had travelled outside the community, and had strong social ties (see Table 1) (Kretzschmar et al., 2013). African American (AA) informants were included in communities where this population exceeded 20% in 1930 (Pederson et al., 1986).

Table 1. DASS speaker types.

Speaker Type	Description
1 "Folk"	Older, less educated, less connected to the community
2 "Common"	Younger, better educated & connected to the community
3 "Cultivated"	Most educated, culturally aware, connected to the community
AA	African American

DASS is a subset of 64 LAGS interviews selected by Lee Pederson to provide a balanced representation of speaker types from each of the 16 regional sectors (Kretzschmar et al., 2013). Four speakers, three European Americans corresponding to each of the speaker types and one African American, were selected for each of the sectors (see Fig. 1). These speakers, 30 female and 34 male, ranged in age from 15 to 90 years old at the time of interviews with a mean age of 61 years old; birth years ranged from 1886 to 1965. Duration of interviews ranged from 2.5 hours to approximately 10 hours with the average interview lasting about 5.75 hours. The formality of interviews varied with some speakers being more talkative and conversational while others were more hesitant and formulaic. In general, however, interviews tended to contain narrative portions, with informants providing continuous speech, and question/answer portions with fieldworkers eliciting lexical items. The 372 hours of audio recordings have been digitized and processed into 405 .wav files ranging from 15 minutes to 1 hour 7 minutes in length, with the average file being 54 minutes.

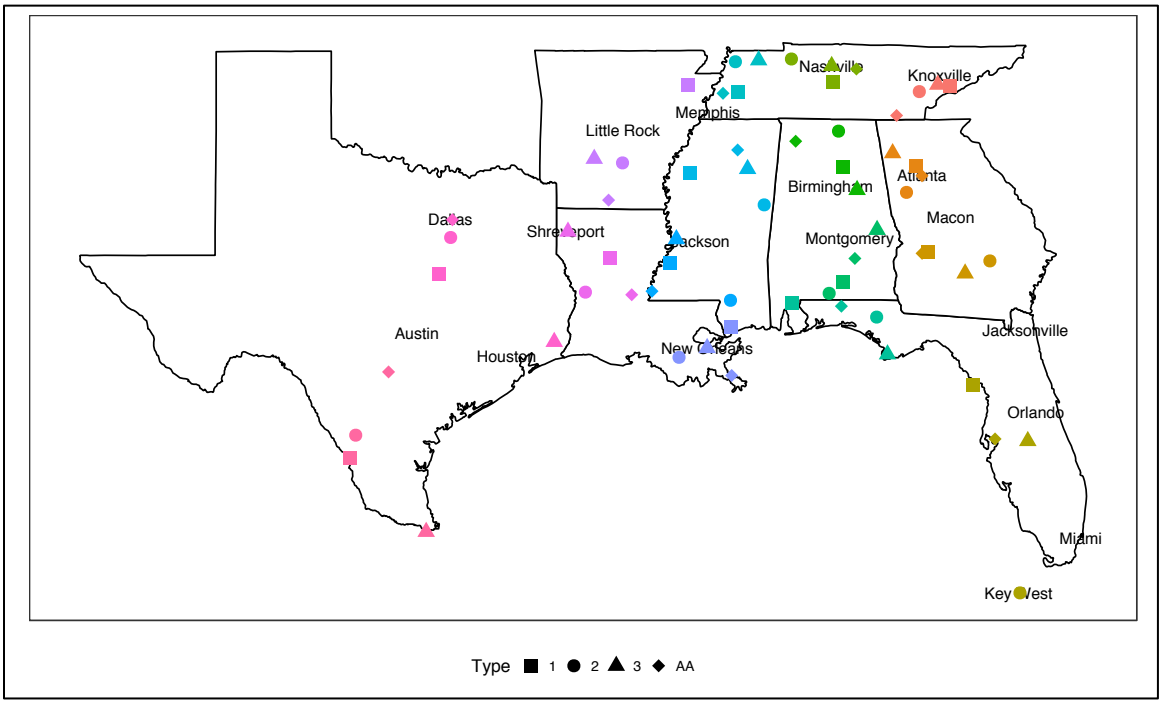


Figure 1. DASS speakers by LAGS sector (color) and type (shape).

B. MOTIVATION FOR TRANSCRIBING DASS

The original goal of LAGS and other Atlas surveys was to document geographic variation and pronunciation of lexical items. Once audio recordings were completed, trained scribes listened and transcribed target words in narrow phonetic script onto speaker Protocols (Pederson, 1981). Figure 2, below, shows an example of one such protocol taken from the portion of the interview where this speaker, an African American male, age 47, from Atlanta, described his childhood home. The protocol contains the impressionistic transcriptions for the target words *bath, kitchen, back porch, bedroom, hall,* and *dining room*. In addition to the target words, brief notes from the scribes describing context was sometimes included in the right-hand margin of the protocols. The original 104-page LAGS survey aimed to collect 1031 target lexical items from a wide range of topics including household descriptions, farm terms, geography, weather terms, and childhood activities, among others.

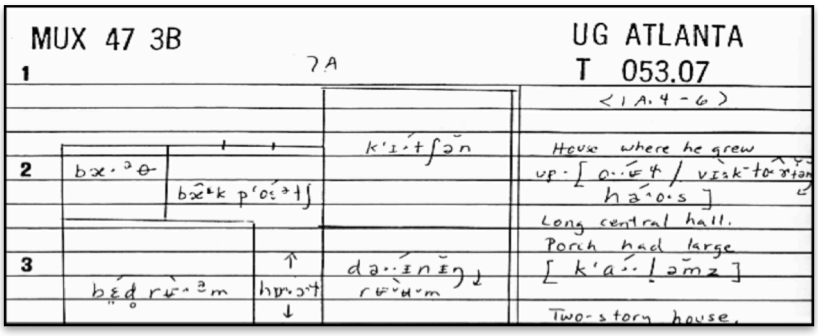


Figure 2. Example of LAGS Speaker Protocol transcriptions.

With such a wealth of data, the Atlas surveys have proven valuable for analyzing lexical and phonological variation in the speech of rural and urban speakers at the respective time periods of collection (Burkette, 2001; Kretzschmar, 2009; Kurath and McDavid, 1961); however, their impressionistic nature and focus on limited target items is quite restricted given what can be achieved with modern methods and technology. Renwick and Olsen's (2016) pilot study of 10 speakers from a single LAGS geographical sector in southeast Georgia illustrated wide variation in speech at the regional and individual speaker levels. Such variation is not reflected in the Protocols, as only representative impressionistic productions were recorded. Furthermore, Renwick and Olsen showed that an exponentially greater number of data points are available through acoustic analysis when full interviews are transcribed. Whereas the LAGS Protocols for the 10 speakers had the potential to show up to 1031 items/speaker (x 10 speakers = 10,310 tokens), the fully transcribed interviews for this small region yielded a corpus of over 132,000 tokens. The present project will transcribe the complete DASS audio corpus (including interviewers' speech, though this is not the focus of the current project) to allow for analysis of all word tokens in the archive. The resulting data set is expected to yield a time-aligned corpus of approximately 1.5 million words. Such a corpus of historical speech is a valuable resource for exploring current models of Southern speech, as well as language change more broadly.

3. METHODS FOR LARGE-SCALE TRANSCRIPTION

A. PREPARING THE AUDIO

The DASS .wav files transcribed in this project are housed at the Linguistic Atlas Project (LAP) (Kretzschmar, 2011), and were digitized from reel-to-reel and cassette recordings. Versions of these files, split into 3-5 minute .mp3 files classified by discussion topic (e.g. geography, dwellings, weather), are publicly available online via the LAP website at www.lap.uga.edu. Given the age of the original recordings and the fact that they were recorded in different years and on different devices, audio quality varies greatly within the corpus. A persistent concern across all .wav recordings was the presence of a high-pitched buzz at approximately 17,000 Hz, visible in the speech signal when the files were viewed in Praat (Boersma and Weenink, 2015). As such, it was necessary to filter the audio. Because some of the .wav files exceed one hour in length, and Praat has difficulty handling such large audio files, a Praat script was run to split each file into six equal pieces, perform a High pass band filter (from 0 to 15,000 Hz, with 100 Hz smoothing), and concatenate the pieces back together.

B. SOFTWARE FOR TRANSCRIPTION

The following computer programs were considered for use in transcribing DASS: An in-house transcription platform called the LAP/LICHEN Transcriber (Kretzschmar, 2011), Praat (Boersma and Weenink, 2015), and Transcriber (Boudahmane et al., 1998). We originally intended to use a specially-designed web-browser-based tool called the LAP/LICHEN Transcriber tool for transcription; however, this transcriber could not generate a TextGrid readable in Praat, and it could not put each speaker on a separate TextGrid tier, both of which are necessary for the purposes of phonetic analysis. Praat (Boersma

and Weenink, 2015) was also considered for its potential to allow for precise transcription according not only to audio cues, but also to visual cues within the speech signal, as well as the ability to transcribe the speech of different speakers into separate tiers; however, given the fact that most of our undergraduate transcribers had no experience with acoustic analysis of speech signals, and the difficulty of working with long sound files in Praat, it was determined that an alternate program should be utilized.

Transcriber (Boudahmane et al., 1998) was eventually chosen as the transcription tool for this project because it offers a solution to each of the aforementioned problems. First, it features a friendly graphical user interface (Fig. 3) that allows for reference to the speech signal (at the level of the waveform), without requiring extensive knowledge of acoustics. Even for a novice to phonetic analysis, it is easy to see that when the waveform is larger, there is probably speech (or any noise that warrants transcription), whereas when the waveform is flat, there is likely silence. This facilitates precise time-alignment of orthographic transcription to the audio signal.

The user opens a .wav file in Transcriber's GUI (Fig. 3), and creates a corresponding Transcriber transcription file (.trs) simply by clicking File > Save as. Tab plays (and pauses) the audio, and as the audio plays, the user enters the corresponding orthographically transcription. When the speaker pauses, or takes a breath, the user presses Enter to start a new chunk (which corresponds to an interval in a Praat TextGrid), denoted in Transcriber as a new line of text starting with a small blue circle (Fig. 3).

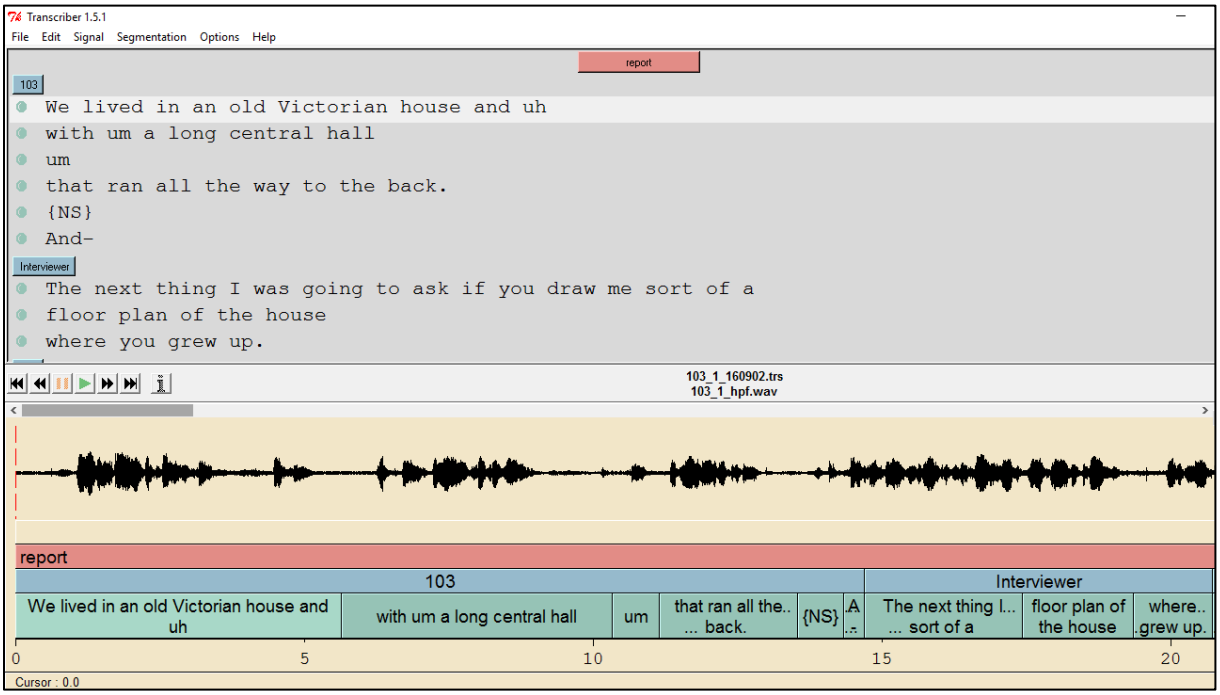


Figure 3. Transcriber software GUI.

The Transcriber GUI allows for easy navigation through the audio and transcription. In Fig. 3 above, the utterance "We lived in an old Victorian house" is highlighted, and the corresponding audio segment is also automatically selected. If the user were to click on the line "that ran all the way to the back", the audio would jump to that section. Similarly, the user can click anywhere in the waveform portion of the GUI, and the corresponding transcription chunk would be highlighted.

C. ENSURING CONSISTENCY

We anticipated it would take undergraduate transcribers 12½ hours of transcription time to complete one hour of audio. Given that DASS consists of 372 hours of audio, it was determined that we would need a large workforce to complete this project. At the time of writing, 48 undergraduate transcribers have contributed to the transcription of DASS. Although an effort is made to assign each transcriber a full

interview (of multiple reels) to allow them to become familiar with both the audio characteristics of the speaker and the audio file, and to maximize transcriber consistency within each interview, this has not always been possible. To mitigate the potential for inconsistency that multiple transcribers can bring, we implemented a program to ensure standardization; below we detail our methods. These include a reference guide of in-house protocols for transcription as well as establishing an in-house dictionary for non-standard words (e.g. *uh-huh, gonna, wanna*), holding an orientation and training session for each new transcriber, and implementing a three-listen system for all files.

i. In-house materials and training

A 19-page Transcription Protocols document was composed with input from previous LAP transcription projects (Kretzschmar, 2011), as well as transcription guidelines produced for the *SLX Corpus of Classic Sociolinguistic Interviews* (Strassel et al., 2003), and recommended by FAVE (Rosenfelder et al., 2014). This document includes basic background on the DASS project, as well as sections on how to use Transcriber (Boudahmane et al., 1998), troubleshoot problems in Transcriber, and save output files. Instructions for, and examples of, creating new conversational turns, punctuation, spelling conventions, contractions, shortened speech, numbers, and miscellaneous sounds are also given. Generally, transcribers are instructed to write exactly what they hear (i.e. it doesn't matter if the transcriber believes an utterance to be "ungrammatical", or a word to be pronounced incorrectly), including interjections, in standard American dictionary English. If a speaker makes a special distinction between how they pronounce a word, and how some other people pronounce that word, this is also noted in the transcription. Standard dictionary spelling is always used, but a comment code is also included on the same line as the alternate pronunciation as shown in Fig. 4 where the second "haunted" production is pronounced with the vowel [eɪ].

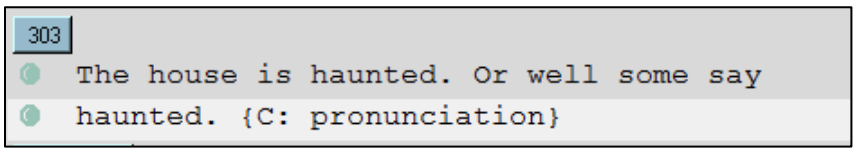


Figure 4. Sample transcription of alternative pronunciation.

A new chunk, or interval, is created whenever a speaker takes a long pause, or when speakers switch (e.g. between Interviewer and primary Speaker). Chunks are prescribed to be not too short and not too long. A transcription of all one-word chunks is unacceptable (as transcribers are not trained to precisely locate word boundaries in the waveform), but we also ask transcribers to avoid chunks that are more than ¾ of a single line in Transcriber in 12-pt type (cf. Fig. 3). Punctuation is limited to capitalization at the beginning of a sentence, and a period or question mark at the end, with sentences often stretching over multiple chunks. Apostrophes are also required for contractions such as *she's*, and hyphens are used for cut-off words (e.g. a speaker may say *Chiff- What is the word? Chifforobe.*). Numbers are always written out, not abbreviated with digits.

For non-standard words, we have created an in-house dictionary that contains our internally agreed-upon spellings for such words as *mm-hmm, um, gonna,* and *wanna.* This dictionary also includes uncommon and antiquated words, often lexical item targets of the initial LAGS survey, that are unfamiliar to the modern transcriber (e.g. *chifforobe*). Students continually expand the dictionary (after consultation with graduate assistants) as they encounter new words, and currently contains over 100 entries in a shared Google spreadsheet. Each entry contains an acceptable spelling of a lexical item (e.g. *gonna*), the accepted spelling of the contracted form of the word (if applicable), and the accepted form of the non-contracted form (if applicable) (*going to*) (Table 2). Transcribers can organize the spreadsheet alphabetically by any of these columns, making potential words and non-standard words easy to find. There is also a fourth column (not shown) for notes, which may include, for instance, definitions of uncommon lexical items.

Table 2. In-house dictionary sample entries.

Lexical Item (or Phrase)	Contracted Form (if applicable)	Non-contracted form (if applicable)
gonna		going to
going to	gonna	

As is the case with most semi-spontaneous speech data, there are multiple instances in each file of non-speech data that must also be recorded. We are attempting to transcribe all audio as quickly and accurately as possible, so we use only a limited number of codes, shown in Table 3. All codes are contained within curly brackets { }.

Table 3. Transcriber codes.

Code	Meaning
{D: }	Doubt: e.g. {D: doubted words}
{X}	Unintelligible
{B}	Beep: previously added by LAP to anonymize audio recordings available to the public
{C: }	Comment: e.g. {C: tape distortion}
{NW}	Non-word: e.g. laugh, cough
{NS}	Non-speech: e.g. dog barking, door closing

A line (i.e. chunk or TextGrid interval) in Transcriber with any curly-bracketed code on it is automatically discarded during forced alignment and subsequent phonetic analysis, although it is retained in the orthographic transcript of the file. Student transcribers therefore separate each curly bracketed code on its own line away from any other transcribed speech (cf. Fig. 3). In this way, we isolate and preserve all phonetically analyzable speech by the main Speaker, even if it occurs directly before or after, for instance, an unintelligible word marked as {X}. In Fig. 4, the second pronunciation of *haunted* is not phonetically analyzed because it has the code {C: pronunciation} on the same line. Transcriber also features a way to mark overlapping speech, which is similarly discarded in our phonetic analysis, but retained in the interview transcript.

Each new undergraduate transcriber participates in a 45-minute to 1-hour long orientation session, conducted by two graduate research graduate assistants on the project. Students are provided with background on LAGS, DASS, and are shown existing LAGS handbooks and materials (Pederson et al., 1986) including original LAGS impressionistic transcription of a LAGS informant. They then receive an abbreviated (1-page) version of the in-house Transcription Protocols that contains basic project information, Transcriber software shortcuts, and the codes shown in Table 3. Students spend approximately 10 minutes working on a sample transcription of the same informant discussed previously in the orientation, during which time guidance on using the software is provided, and students begin to familiarize themselves with both the software and the transcription procedures. If a student is hired to work on the project, they receive access to a Google drive folder containing the following: Transcription Protocols, the abbreviated 1-page Protocols, a 1-page document of tips compiled from more experienced transcribers, the in-house phrase dictionary, and a spreadsheet log on which students track their transcription progress.

ii. Three-listen system

Each transcriber is assigned one reel of audio at a time, and transcribers are responsible for listening to that reel twice. In the first listen, the transcriber is generally attempting to record who said what and when (e.g. the main Speakers vs. Interviewer vs. any Auxiliary speakers). In the second listen, the transcriber ensures that chunks are correctly time-aligned to the audio, overlapping speech is correctly marked, and that punctuation, spelling, and coding conventions have been appropriately followed. We have also found that listening to the audio all the way through twice aids transcribers in understanding difficult audio that they may not have understood the first time around.

Once a student transcriber has completed two listens of the audio, they alert the project graduate assistants of the reel's completion through a shared spreadsheet time log. A graduate student then performs a third listen of the audio. During this final listen we bring transcriptions into greater consistency with one another. In this third listen, graduate assistants are particularly checking that all main speaker speech is chunked appropriately, that all speech has been recorded accurately, and that in-house conventions have been followed. Common errors include codes not separated onto their own lines, imprecisely time-aligned chunk boundaries, and spelling inconsistencies. The graduate assistants make corrections to the files as necessary.

D. FILE STORAGE AND TRACKING

Given the amount of labor it takes to transcribe the audio, we have taken steps to ensure that no data is lost due to technical malfunction. The LAP office has eight desktop computers, and each transcriber is assigned a specific computer and time window (typically up to 10 hours a week) in which to work. On the assigned desktop, the transcriber has a personal folder in which .wav audio and .trs Transcriber files are stored. The Transcriber files are date-coded as follows: 103_1_160902.trs (as shown in Fig. 5) where 103 = the DASS-assigned speaker number, 1 = reel number, 16 = two-digit year (2016), 09 = two-digit month (September), and 02 = two-digit day (the 2nd). Each day that a student transcribes, they open the previous transcription session's .trs file and immediately save it as a new date-coded file. Using the example above, if a student transcribed two days later, they would name their new file 103_1_160904.trs. By the end of a full file, an undergraduate transcriber will have multiple versions of the same file in their personal Desktop folder, each building on the last (Fig. 5). When a transcriber has finished transcribing a file, the date-coded filename is appended with _final.trs.

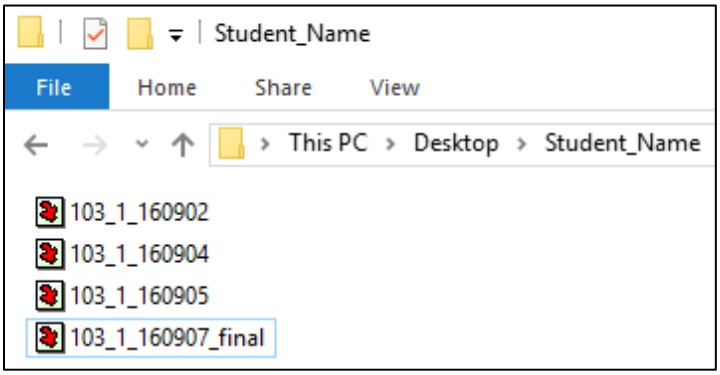


Figure 5. Successive date-coded .trs files.

Each computer is also connected to an external hard drive, on which each transcriber has their own folder. After each transcription session, they save a copy of the most recent date-coded file in that folder, so that they essentially have two identical personal folders. Graduate research assistants perform weekly backups of all work to the internal office server.

As student transcribers get comfortable with using the Transcriber software, and with transcribing the DASS audio in general, they are offered the opportunity to work remotely. This option is usually offered, though not required, after students have completed at least two acceptable reels. Those who transition out of the office receive a USB drive on which typically two reels of audio are loaded. They are advised to save copies of their date-coded transcriptions on their personal computers, as well as on the provided USB drive. At least twice monthly, these students come into the office to check in at which time graduate assistants back up any new transcriptions on the LAP main server, and provide the student with more audio if necessary. Given the limited LAP office space and number of computers, it is only by cycling transcribers out of the physical office that we have been able to build our transcription workforce.

In order to keep track of which transcribers are working on which files, and how far into the transcription the transcriber has progressed, students fill out a spreadsheet time log in our shared Google drive folder after each transcription session as shown in Fig. 6.

C	D	E	F	G	H	I	J	K
Informant	Reel	State	Time	Time	Transcriber	First Listen	Second Listen	Back-Up
INF185	1	GA	0:59:11	0:59:11	JK	DONE 11/02/16	DONE 11/11/16	RMO 11/18/2016
INF185	2	GA	0:58:53	0:58:53	JK	DONE 1/18/17	DONE 1/25/17	RMO 2/14/2017
INF185	3	GA	0:58:33	0:58:33	JK	DONE 2/13/17	DONE 2/15/17	RMO 2/21/2017
INF185	4	GA	0:58:27	0:58:27	JK	DONE 4/17/17	DONE 4/19/17	RMO 4/25/2017
INF185	5	GA	0:58:47	0:58:47	JK	4/19/17 1m29s		RMO 4/25/2017

Figure 6. Transcription time log.

The spreadsheet is organized by DASS-assigned speaker number, then reel. When audio is provided to a transcriber, that transcriber's initials are written in column H on the spreadsheet. As the transcription progresses, the transcriber updates the spreadsheet after each session with the most recent date of transcription and with how many minutes (m) and seconds (s) into the transcription the transcriber has gotten. This process is first done for the first listen (column I), and then the second listen (column J). In Fig. 6 above, transcriber JK has completed four reels of speaker 185's interview. JK is currently 1 minute and 29 seconds into the first listen of speaker 185's reel 5. In column K, graduate assistants initial and date to track the backup of these transcriptions to the main LAP server.

4. WORKFLOW

At the time of writing, 19 complete interviews have been transcribed, and at least some audio from 53 of 64 interviews has been transcribed, in a period of approximately 9 months. 201 complete reels (of 405 total) have been transcribed yielding over 179 hours of audio. For 44 of these interviews, at least one reel of audio has also been processed and force-aligned (totaling over 142 hours of audio processed thus far). The workflow for transcribing, force-aligning, and processing this data is as follows:

- Transcription by undergraduate using Transcriber, including double-checking
  - Final product is appended with _final.trs
- Spot-checked for consistency in Transcriber by graduate assistant (3rd listen)
  - Appended with _final_checked.trs
- File conversion via LaBB-CAT (Fromont and Hay, 2012) scripts
  - .trs (.xml) → .txt
  - .trs (.xml) → .TextGrid
- Automatic phonetic analysis (see §5)

The faculty and graduate students working on the project track completed reels in a shared, remotely accessible spreadsheet as shown in Fig. 7.

C	D	F	G	H	I	K	M	N
Informant	Reel	Time	Transcriber	Transcription Complete (_final.trs)	3rd Listen (_final_checked.trs)	prepareFor DARLA.praat	DARLA-ed and post-processed	Notes
INF027	1	1:00:15	JC	09/30/2016	9/30/2016	12/6/2016	12/6/2016	
INF040	1	0:44:36	SK	10/4/2016	10/4/2016	10/11/2016	10/13/2016	
INF040	2	0:44:26	SK	10/7/2016	10/7/2016	10/14/2016	10/14/2016	

Figure 7. Spreadsheet for tracking processing and forced alignment of completed reels.

Time alignments from Transcriber are mapped to TextGrid intervals, so that each chunk is its own interval, and each speaker receives a separate tier, as shown in Fig. 8. The interviewer is Int103 (tier 1),

the main speaker is 103 (tier 2), and other speech (e.g. curly bracketed codes such as non-speech {NS}) is marked as Other (tier 3). Only the main speaker tier (2) is phonetically analyzed.

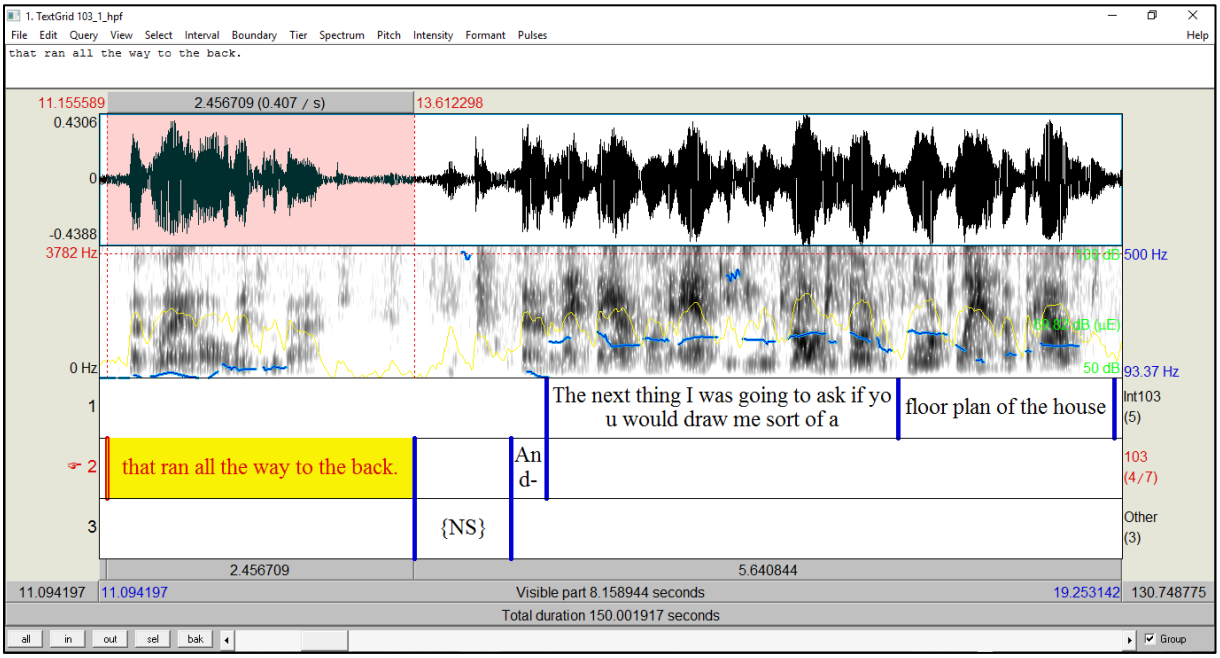


Figure 8. Praat TextGrid resulting from Transcriber file.

5. FORCED ALIGNMENT AND FORMANT EXTRACTION

TextGrids and accompanying .wav files are sent to the Dartmouth Linguistic Automation suite (DARLA) (Reddy and Stanford, 2015), which uses ProsodyLab Aligner (Gorman et al., 2011) for forced alignment, and performs extraction of acoustic data using a version of FAVE-Extract (Rosenfelder et al., 2014). A sample force-aligned TextGrid is shown in Fig. 9. DARLA also returns vowel formant measurements; however, it currently also automatically filters data based on vowel duration, bandwidth, and other factors, and by default does not return measurements for every token. The main drawback of this is that it potentially filters out speech variation in which we are interested, and we have no control of which tokens are filtered.

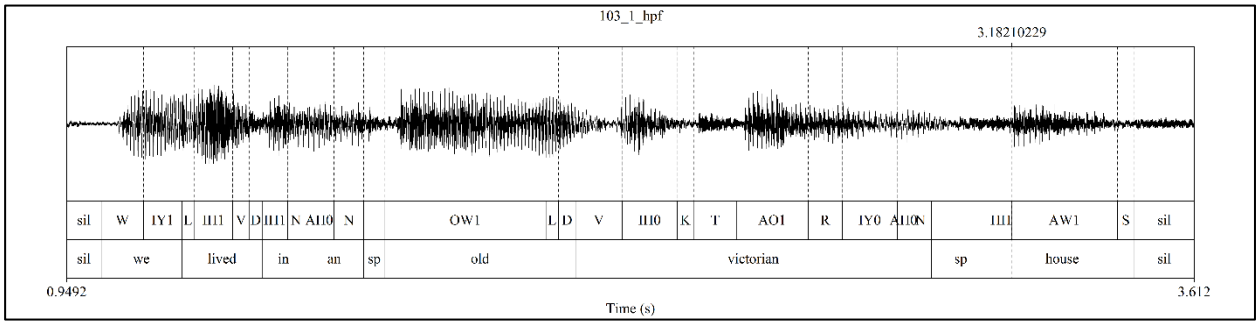


Figure 9. Force-aligned TextGrid returned by DARLA.

We are currently testing DARLA against three non-filtering formant extraction techniques: an in-house Praat (Boersma and Weenink, 2015) script, the Forced Alignment and Vowel Extraction program (FAVE-Extract only) (Rosenfelder et al., 2014), and a modified version of FAVE-Extract trained on our own

DASS data. Results of these formant extractors on DASS speaker 556 (a 77-year-old European American male from Mississippi) are shown in Fig. 10.

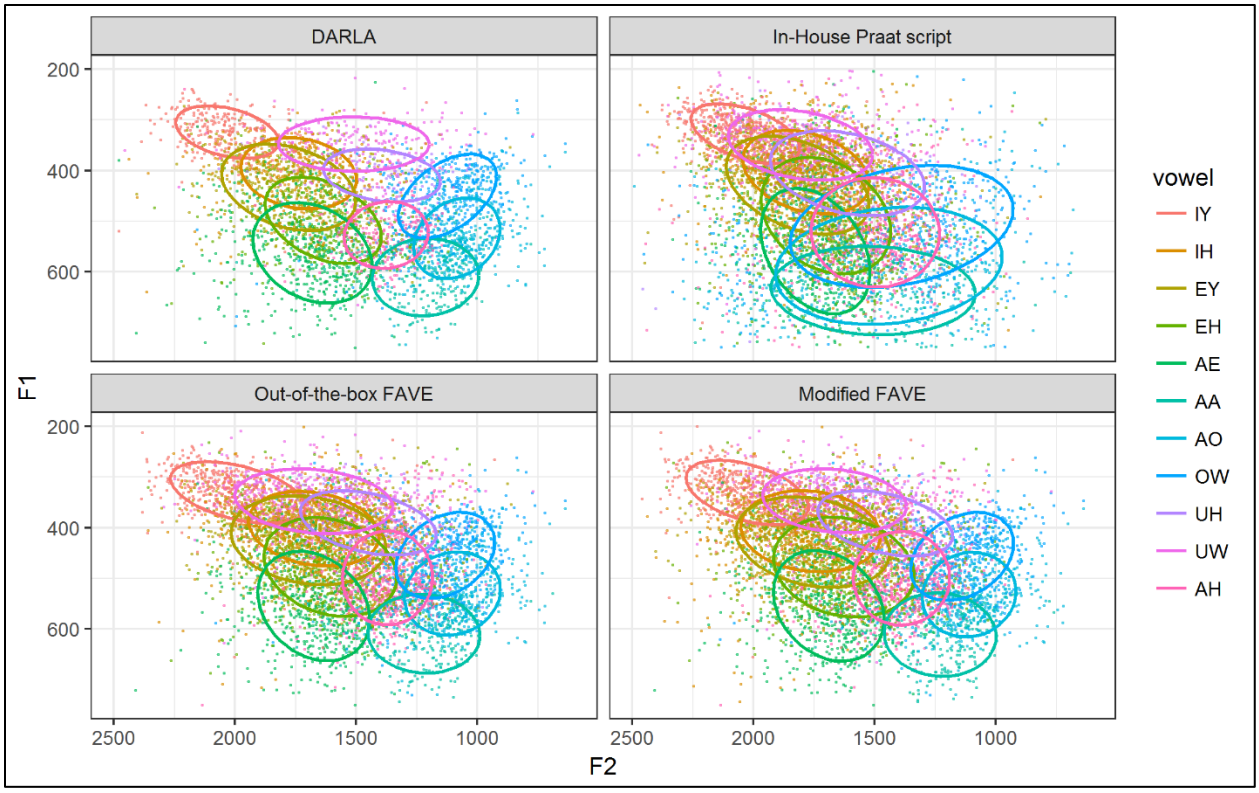


Figure 10. Comparison of vowel formant extraction methods; ellipses (1SD) by vowel.

The DARLA formant extraction (top left) above shows a clearly less populated vowel space, and in fact, because of our noisy audio, DARLA has filtered out 53% of potential data points. Our in-house Praat script (top right) is successful at extracting all data, as evidenced by the much more densely populated vowel space, but the severe overlap of all vowels indicates that the formant tracking is prone to error, particularly for back vowels, whose low second-formant measurements may be mistaken for the first formant measurement, or missed, by Praat. The out-of-the-box version of FAVE-Extract (bottom left) extracts all tokens, but its training data come from many U.S. varieties from the *Atlas of North American English* (Labov et al., 2006), and its Bayesian formant tracking algorithm is not specialized for Southern speech (cf. Labov et al., 2013). The in-house modified version of FAVE's (bottom right) Bayesian algorithm is trained on mean formant values and their distribution from 4 fully transcribed DASS interviews (Table 4).

Table 4. Speakers used for Modified FAVE training data.

Speaker	Age	Gender	Ethnicity	State
40	38	F	EA	TN
434	90	M	AA	AL
657X	86	M	EA	LA
894	74	M	EA	TX

Of these four formant extraction methods, modified FAVE appears to perform best. Although it requires an extra step for data extraction (with respect to DARLA, which does alignment and extraction together), it provides a clean, well-separated vowel space similar to that produced by DARLA, but

without data loss due to filtering. As we acquire more completed speakers from more diverse demographic backgrounds, we plan to add more speakers to the training data in the hopes of even more accurately extracting formant values.

6. CONCLUSION

Historical speech corpora are valuable sources of acoustic data that may inform current models of speech and language change. As an extensive and well-sampled corpus of Southern speech, DASS has much to offer to modern sociophonetic research; however, as it previously lacked a full transcription, it was not possible to apply methods of large-scale phonetic analysis to the corpus. Here we have laid out our step-by-step methods for transcribing and force-aligning DASS, as well as for managing the large amount of data produced in this pursuit. We began the process with unfiltered .wav files and impressionistic transcriptions in the LAGS Protocols of a limited number of target lexical items. In creating and force-aligning this corpus, the following resources are now available:

- Filtered .wav files of DASS recordings
- Transcriptions in .trs (xml), .txt format (with all speakers in file)
- TextGrids whose annotations are aligned to the audio at the segment level
- For each file sent to DARLA, a spreadsheet of acoustic values with many associated annotations including phonological characteristics for each vowel token, speaker characteristics, etc. (for main speaker only)
- Spreadsheets of acoustic values from the other three extraction methods

Ultimately, our time-aligned orthographic companion to the DASS audio corpus allows for automatic phonetic analysis not previously possible on these recordings. Acoustic analysis results are available in Renwick and Stanley (2017), and interested readers may interact with the dataset and find additional resources related to DASS via the Gazetteer of Southern Vowels (Stanley et al., 2017). We hope that the methods described herein are helpful to others with access to large audio corpora, and that interest may be sparked in fully utilizing acoustic data already in hand.

ACKNOWLEDGMENTS

This research is supported by NSF BCS #1625680 awarded to co-PIs Kretzschmar and Renwick, the University of Georgia Graduate School, and the American Dialect Society. We are grateful to Joe Fruehwald for instructions on how to modify FAVE.

REFERENCES

Boersma, P., and Weenink, D. (**2015**). Praat: Doing phonetics by computer [Computer program], Version 5.4.08. Retrieved from http://www.praat.org

Boudahmane, K., Manta, M., Antoine, F., Galliano, S., and Barras, C. (**1998**). Transcriber: a tool for segmenting, labeling and transcribing speech [Computer program], Version 1.5.2. Retrieved from http://trans.sourceforge.net/

Burkette, A. (**2001**). "The story of chester drawers," Am. Speech, **76**, 139–157.

Coleman, J., Baghai-Ravary, L., Pybus, J., and Grau, S. (**2012**). *Audio BNC: The Audio Edition of the Spoken British National Corpus*, Available: http://www.phon.ox.ac.uk/AudioBNC

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., and Martey, N. (**2000**). *Santa Barbara Corpus of Spoken American English, Parts 1-4,* Linguistic Data Consortium, Available: http://www.linguistics.ucsb.edu/research/santa-barbara-corpus

Fromont, R., and Hay, J. B. (**2012**). "LaBB-CAT," Proc. Australas. Lang. Technol. Workshop, 113–117.

Gorman, K., Howell, J., and Wagner, M. (**2011**). "Prosodylab-Aligner: A tool for forced alignment of laboratory speech," Can. Acoust., **39**, 192–193.

Kretzschmar, W. A. (**1994**). *Handbook of the Linguistic Atlas of the Middle and South Atlantic States.*, University of Chicago Press, Chicago, 454 pp.

Kretzschmar, W. A. (**2009**). *The Linguistics of Speech*, Cambridge University Press, Cambridge; New York, 296 pp.

Kretzschmar, W. A. (**2011**). *Linguistic Atlas Project*, Linguist. Atlas Proj., Available: http://www.lap.uga.edu
Kretzschmar, W. A., Bounds, P., Hettel, J., Pederson, L., Juuso, I., Opas-Hänninen, L. L., and Seppänen, T. (**2013**). "The Digital Archive of Southern Speech (DASS)," South. J. Linguist., **37**, 17–38.
Kurath, H., Hansen, M. L., and Bloch, B. (**1939**). *Linguistic Atlas of New England*, Brown University, Providence, RI, 240 pp.
Kurath, H., and McDavid, R. I. (**1961**). *The Pronunciation of English in the Atlantic States: Based upon the Collections of the Linguistic Atlas of the Eastern United States*, University of Michigan Press, Ann Arbor, MI, 182 pp.
Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology, and Sound Change: A Multimedia Reference Tool*, Mouton de Gruyter, Berlin; New York, 318 pp.
Labov, W., Rosenfelder, I., and Fruehwald, J. (**2013**). "One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis," Language, **89**, 30–65.
Pederson, L. (**1981**). *Protocol of a Field Record for the Linguistic Atlas of the Gulf States*, University Microfilms International, Ann Arbor, MI.
Pederson, L., McDaniel, S. L., and Adams, C. M. (Eds.) (**1986**). *Linguistic Atlas of the Gulf States*, University of Georgia Press, Athens, GA, Vols. 1-7.
Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (**2005**). "The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability," Speech Commun., **45**, 89–95. doi:10.1016/j.specom.2004.09.001
Reddy, S., and Stanford, J. N. (**2015**). "Toward completely automated vowel extraction: Introducing DARLA," Linguist. Vanguard, **1**, 15–28.
Renwick, M. E. L., and Olsen, R. M. (**2016**). "Voices of coastal Georgia," Proc. Meet. Acoust., **25**, 060004. doi:10.1121/2.0000176
Renwick, M. E. L., and Stanley, J. A. (**2017**). "An acoustic perspective on legacy data: Vowels in the Digital Archive of Southern Speech," J. Acoust. Soc. Am., **141**, 3981–3981. doi:http://dx.doi.org/10.1121/1.4989091
Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gormon, K., Prichard, H., and Yuan, J. (**2014**). *FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2*, Available: 10.5281/zenodo.22281
Stanley, J. A., Kretzschmar, W. A., Renwick, M. E. L., Olsen, M. L., and Olsen, R. M. (**2017**). *Gazetteer of Southern Vowels*, Available: http://lap3.libs.uga.edu/u/jstanley/vowelcharts/
Strassel, S., Conn, J., Wagner, S. E., Cieri, C., Labov, W., and Maeda, K. **(2003)**. "Automatic Alignment and Analysis of Linguistic Change - Transcription Guidelines," Linguistic Data Consortium. Retrieved from http://projects.ldc.upenn.edu/DASL/SLX/docs/transcription.pdf