JDHCS 2010
Volume 1 Number 2

Developing Web Interfaces to Spoken Language Data Collections

Tyler Kendall, Linguistics Department, Northwestern University

Introduction

Across disciplines, and particularly in linguistics, there have been major movements towards the digitization of older audio recordings in recent years, and current recordings are now most often "born-digital". At the same time, powerful and freely available tools are increasingly available for analyzing these resources, such as Praat,[1] the open-source phonetics software, ELAN,[2] an open-source tool for linguistic annotation of audio and video data, and R,[3] the open-source statistical computing language and environment. Together, these factors—the ubiquity of digital recordings and the wide availability of software tools—have led to an environment where powerful analysis of speech recordings and the development of large collections of audio-based data are possible with few resources beyond a personal computer. This has, without doubt, enabled new approaches and wider access to the analysis of spoken language data across disciplines, especially within linguistics.

While there has been great focus on the benefits of the new, digital audio technologies, within academia less attention has been paid to the difficulties of keeping track of these digital recordings and to their maintenance and management. That is, the long-term preservation of these materials is not straightforward, and, if not properly cared for, digital recordings may be easier to lose or lose access to than their analog counterparts.[4] A consideration of the best methods of organization and preservation for these digital resources leads in turn to a fuller understanding of the great possibilities that current technologies enable in terms of our interfaces with our data and the ways that we present, interact with, and analyze audio-based language data.[5]

In this short paper, I exemplify some possibilities for web-based databases of audio, spoken language data by briefly discussing two related projects, the Sociolinguistic Archive and Analysis Project[6] and the Online Speech/Corpora Archive and Analysis Resource.[7]

The Sociolinguistic Archive and Analysis Project (SLAAP)

SLAAP, which has been outlined elsewhere,[8] is a web-based digitization and preservation project housed at North Carolina State University, featuring a growing archive of sociolinguistic audio

[1] http://www.fon.hum.uva.nl/praat/; Boersma and Weenink 2009.

[2] http://www.lat-mpi.eu/tools/elan/

[3] http://www.r-project.org; R Development Core Team 2009.

[4] cf. Bird and Simons 2003.

[5] cf. Kendall 2008.

[6] SLAAP; http://ncslaap.lib.ncsu.edu.

[7] OSCAAR; http://oscaar.ling.northwestern.edu.

[8] e.g., Kendall 2007, 2008, 2009; Newman 2008.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

recordings along with dynamic interfaces to those recordings. At the time of this writing, over 1,600 interview recordings are stored in and accessible through SLAAP, amounting to over 1,275 hours of recorded speech. The centerpiece of the SLAAP software is a time-aligned annotation framework that is integrated with analytic software (including Praat and R) allowing for features like the automatic generation of spectrograms within the web-based audio player, the extraction of phonetic data from within a recording's transcript, multiple and dynamic displays of each transcript, and corpus linguistic analyses across the diverse materials in the archive. At present (only) 3% of the total archive is transcribed. This constitutes a 370,000 word searchable corpus, representing 37 hours of recorded talk, time-aligned at the utterance level. Like the digitization and entry of the audio recordings, transcription is ongoing, though—as is evident from the small percentage of transcribed talk—it is more slow-going than digitization. Nonetheless, SLAAP and its time-aligned transcript collection have proven useful for a wide-range of research and educational uses.[9] (The transcript model, method, and conventions are spelled out in Kendall 2007 and 2009, and in the SLAAP User Guide available at http://ncslaap.lib.ncsu.edu/userguide/. Several screenshots and further discussions of the software and archive are also available in those sources.)

The Online Speech/Corpora Archive and Analysis Resource (OSCAAR)

OSCAAR, a more recent initiative, just begun in the Fall of 2009 in the Speech Communication Research Lab of the Northwestern University Linguistics Department, represents an attempt to extend SLAAP's data model and analysis features to a broader audience of users. While SLAAP has been designed specifically around sociolinguistic applications and datasets (e.g., in terms of metadata elements, annotation schemes, and user-features), OSCAAR is designed to be more generically useful. It seeks to provide a web-based storage and access facility for speech recordings of all sorts, and features for their analysis, with fewer assumptions about the kinds of recordings that will be housed on the website or the needs and interests of its users. OSCAAR allows for the flexible organization of various kinds of speech recordings so that researchers can more easily

> (a) retrieve and review possible speech recordings for acoustic or other linguistic analysis (as in Figure 1);
> (b) retrieve and review possible speech recordings for use as stimuli in speech recognition and perception experiments;
> (c) gain new perspectives on their data (see Figure 4 for an example); and, of course,
> (d) preserve recording collections, and share them, via the web-interface, with colleagues.

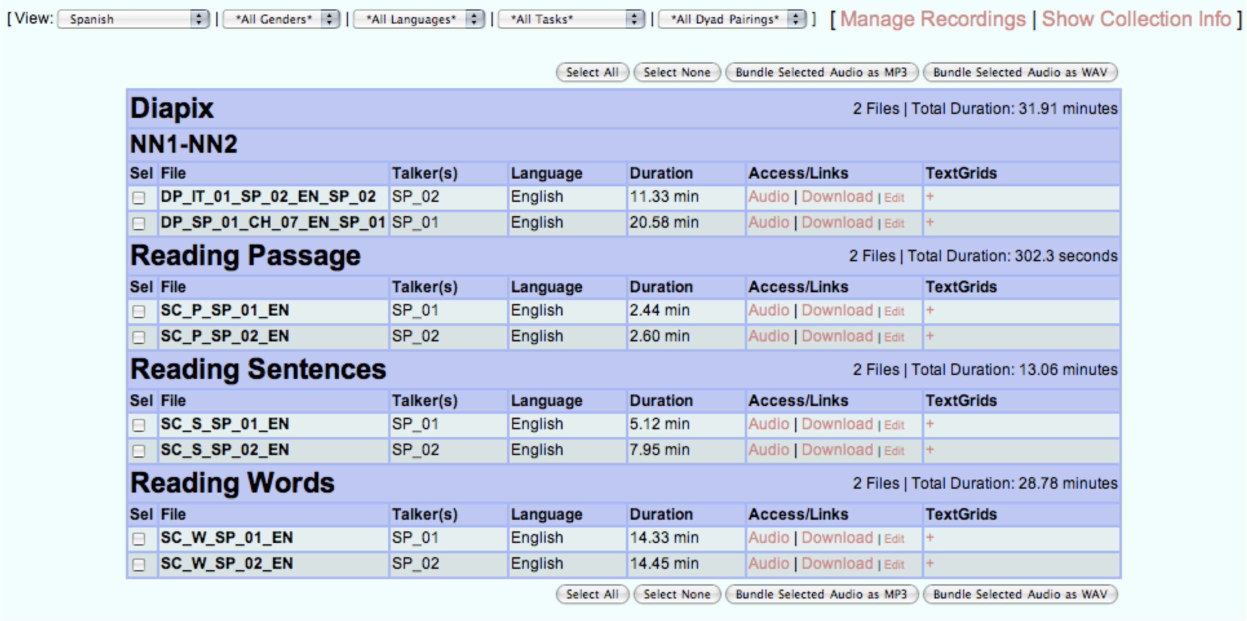
While still in its early phases of development, OSCAAR already houses many of the speech recording collections developed by the Speech Communication Research Lab and we are beginning to add data collections from other researchers. At the time of this writing, there are approximately 4,500 recordings stored in OSCAAR. These are primarily recorded sentences used (and useable) as stimuli for speech perception experiments, but other task-based recordings are also available in the archive. As an example of OSCAAR's interface, Figure 1 displays a screenshot of English language recordings in the Wildcat Corpus[10] from native Spanish language talkers. Users can view, explore, and search the recordings in a variety of other ways, from characteristics of the talker, to collection-specific organizational terms (the "Diapix", "NN1-NN2", "Reading Passage", "Reading Sentences",

[9] e.g., Carter 2009; Kendall 2009; Kendall, Van Herk, and Bresnan 2009; Kohn 2008; Mallinson and Kendall 2009.

[10] Van Engen, Baese-Berk, Baker, Choi, Kim, and Bradlow *in press.*

Published by: The Division of the Humanities at the University of Chicago

and "Reading Words" headers in Figure 1 are all custom, sort-able organizational terms within the Wildcat Corpus; other collections can use other terms and categories), and, finally, to complex searches across multiple collections.


[View: Spanish | *All Genders* | *All Languages* | *All Tasks* | *All Dyad Pairings*] [Manage Recordings | Show Collection Info]

Diapix 2 Files | Total Duration: 31.91 minutes

NN1-NN2

Sel	File	Talker(s)	Language	Duration	Access/Links	TextGrids
□	DP_IT_01_SP_02_EN_SP_02	SP_02	English	11.33 min	Audio | Download | Edit	+
□	DP_SP_01_CH_07_EN_SP_01	SP_01	English	20.58 min	Audio | Download | Edit	+

Reading Passage 2 Files | Total Duration: 302.3 seconds

Sel	File	Talker(s)	Language	Duration	Access/Links	TextGrids
□	SC_P_SP_01_EN	SP_01	English	2.44 min	Audio | Download | Edit	+
□	SC_P_SP_02_EN	SP_02	English	2.60 min	Audio | Download | Edit	+

Reading Sentences 2 Files | Total Duration: 13.06 minutes

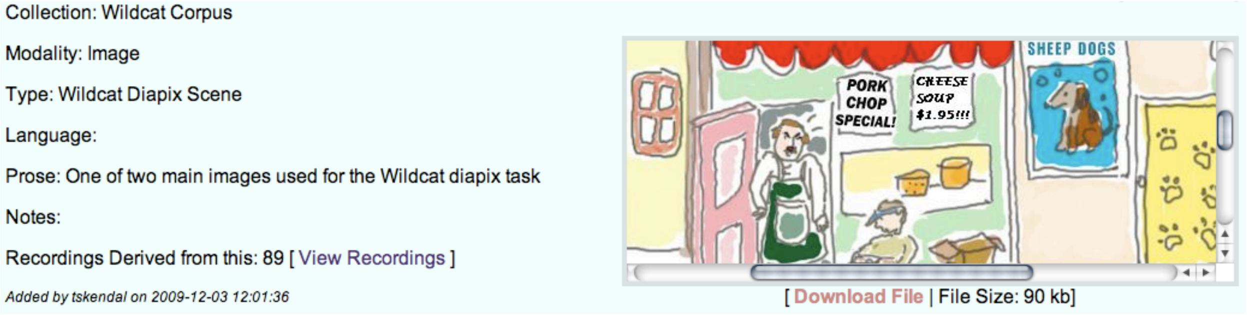
Sel	File	Talker(s)	Language	Duration	Access/Links	TextGrids
□	SC_S_SP_01_EN	SP_01	English	5.12 min	Audio | Download | Edit	+
□	SC_S_SP_02_EN	SP_02	English	7.95 min	Audio | Download | Edit	+

Reading Words 2 Files | Total Duration: 28.78 minutes

Sel	File	Talker(s)	Language	Duration	Access/Links	TextGrids
□	SC_W_SP_01_EN	SP_01	English	14.33 min	Audio | Download | Edit	+
□	SC_W_SP_02_EN	SP_02	English	14.45 min	Audio | Download | Edit	+

Figure 1. OSCAAR interface to a subset of recordings for the Wildcat Corpus

In addition to speech recordings themselves, OSCAAR also stores information about the talkers, and the associated materials for those recordings, such as the reading passages, interview protocols, or other stimuli used in eliciting the speech. These can be reviewed within OSCAAR and users can quickly peruse the recordings derived from a given material. This is demonstrated in Figure 2, a picture used for the "Diapix" task in the Wildcat Corpus.[11]


Collection: Wildcat Corpus

Modality: Image

Type: Wildcat Diapix Scene

Language:

Prose: One of two main images used for the Wildcat diapix task

Notes:

Recordings Derived from this: 89 [View Recordings]

Added by tskendal on 2009-12-03 12:01:36

[Download File | File Size: 90 kb]

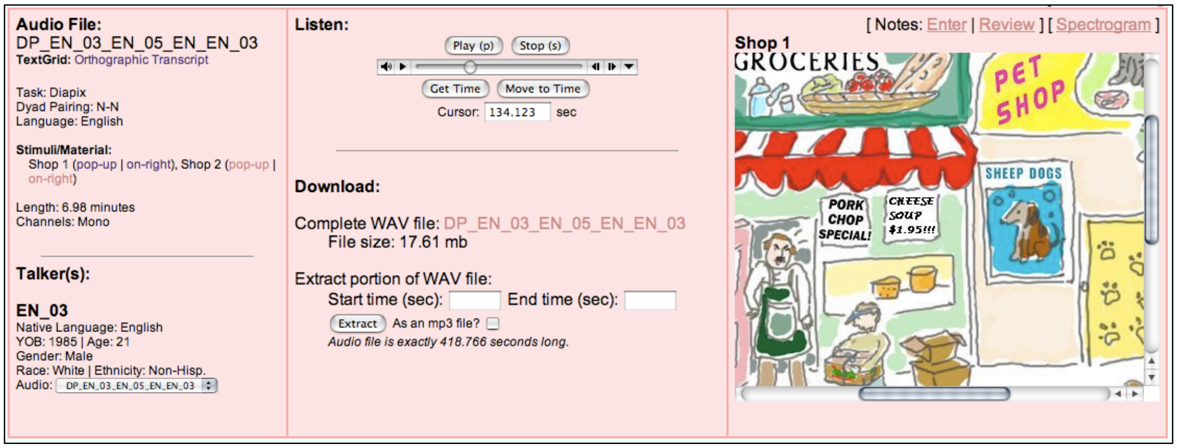
Figure 2. OSCAAR interface to one of the picture stimuli for the Wildcat Corpus

Figure 3 displays an inline audio player for one of the Wildcat Corpus recordings derived from the picture shown in Figure 2. Here a user can review the stimuli (in this case a picture) while listening to

[11] Van Engen et al. *in press.*

the audio. Not shown in Figure 3 are other "value-added" features of the audio player, such as the ability to create—and later search and/or return to—time-stamped notes that can be associated to specific moments in an audio recording. Users can also download complete audio files or extract portions of the audio by entering a time-range.

Additionally, OSCAAR provides a static URL and a "homepage" (not shown) for each data collection, so that researchers can publish persistent links to their dataset and then chose how much access to give public viewers (visitors to OSCAAR can view basic information about data collections, but must be granted access by the site's administrator and the collection's owner to access the actual resources).



Audio File:
DP_EN_03_EN_05_EN_EN_03
TextGrid: Orthographic Transcript

Task: Diapix
Dyad Pairing: N-N
Language: English

Stimuli/Material:
Shop 1 (pop-up | on-right), Shop 2 (pop-up | on-right)

Length: 6.98 minutes
Channels: Mono

Talker(s):

EN_03
Native Language: English
YOB: 1985 | Age: 21
Gender: Male
Race: White | Ethnicity: Non-Hisp.
Audio: DP_EN_03_EN_05_EN_EN_03

Listen:

Cursor: 134.123 sec

Download:

Complete WAV file: DP_EN_03_EN_05_EN_EN_03
File size: 17.61 mb

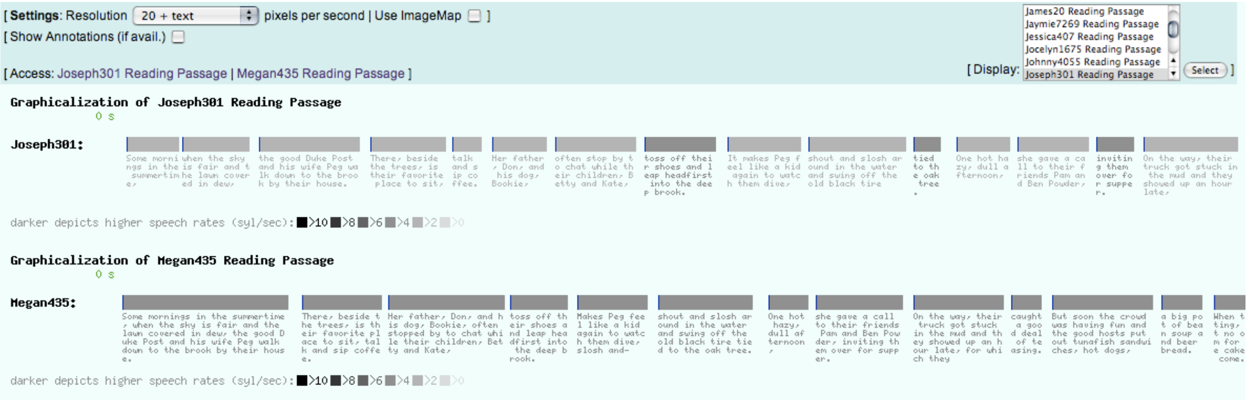
Extract portion of WAV file:
Start time (sec): End time (sec):
As an mp3 file? □
Audio file is exactly 418.766 seconds long.

[Notes: Enter | Review] [Spectrogram]

Shop 1

Figure 3. Audio access within OSCAAR, showing an inline view of the picture about which the recorded talk (in this case) centers.

Building on some of the features developed for SLAAP, OSCAAR provides a range of advanced features that provide new insight into spoken language recording collections. For instance Figure 4, using data from Kendall and Fridland,[12] provides one example of a way that OSCAAR dynamically provides new ways to interface with and view speech data.



[**Settings**: Resolution 20 + text pixels per second | Use ImageMap □]
[Show Annotations (if avail.) □

[Access: Joseph301 Reading Passage | Megan435 Reading Passage]

[Display: James20 Reading Passage
Jaymie7269 Reading Passage
Jessica407 Reading Passage
Jocelyn1675 Reading Passage
Johnny4055 Reading Passage
Joseph301 Reading Passage]

Graphicalization of Joseph301 Reading Passage
0 s

Joseph301:
Some mornings in the summertime,
when the sky is fair and the lawn covered in dew,
the good Duke Post and his wife Peg walk down to the brook by their house.
There, beside the trees, is their favorite place to sit,
talk and sip coffee.
Her father, Don, and his dog, Bookie,
often stop by to chat while their children, Betty and Kate,
toss off their shoes and leap headfirst into the deep brook.
It makes Peg feel like a kid again to watch them dive,
shout and slosh around in the water and swing off the old black tire
tied to the oak tree.
One hot hazy, dull afternoon,
she gave a call to their friends Pam and Ben Powder,
inviting them over for supper.
On the way, their truck got stuck in the mud and they showed up an hour late,

darker depicts higher speech rates (syl/sec): ■>10 ■>8 ■>6 ■>4 ■>2 ■>0

Graphicalization of Megan435 Reading Passage
0 s

Megan435:
Some mornings in the summertime, when the sky is fair and the lawn covered in dew, the good Duke Post and his wife Peg walk down to the brook by their house.
There, beside the trees, is their favorite place to sit, talk and sip coffee.
Her father, Don, and his dog, Bookie, often stopped by to chat while their children, Betty and Kate,
toss off their shoes and leap headfirst into the deep brook.
Makes Peg feel like a kid again to watch them dive, slosh and-
shout and slosh around in the water and swing off the old black tire tied to the oak tree.
One hot hazy, dull afternoon,
she gave a call to their friends Pam and Ben Powder, inviting them over for supper.
On the way, their truck got stuck in the mud and they showed up an hour late, for which they
caught a good deal of teasing.
But soon the crowd was having fun and the good hosts put out tunafish sandwiches, hot dogs,
a big pot of bean soup and beer bread.
When t ting, t no o m for e cake come.

darker depicts higher speech rates (syl/sec): ■>10 ■>8 ■>6 ■>4 ■>2 ■>0

[12] Kendall and Fridland 2010.

Figure 4. OSCAAR *graphicalization* of aligned orthographic text for two talkers reading the same reading passage

This *graphicalization*[13] depicts the speech timing of talkers reading (aloud) the same reading passage. The darkness of shading of each gray block indicates the rate of speech (automatically determined by the software) and the width of each block its duration. Blank areas reflect silent pauses, with the width of the blank area depicting the duration of the silent pause. The version displayed in Figure 4 is set to a resolution of 20 pixels per second. The text of each utterance is displayed beneath each block. (The image scrolls off-screen to the right.) Through this view, we readily see, for example, that the first talker, Joseph301, reads the passage more slowly and with more pauses than the second talker, Megan435. This is confirmed by a quantitative assessment of the data. (Joseph301 has a median articulation rate of 3.8 syllables per second while Megan435 has a median articulation rate of 4.8 syllables per second. Joseph301 also realizes 28 pauses during the reading, compared to only 17 by Megan435.)

Conclusion

SLAAP has enabled researchers to interact with and investigate their data in new ways.[14] OSCAAR, we believe, will allow a wider-range of researchers to benefit from these software features and, at the same time, provides an opportunity to revisit some of SLAAP's original questions: How can we not only preserve but best manage large and growing collections of audio data? How can we leverage new technologies so that our data archives are not just useable, but maximally useful?

Overall, we argue that web-enhanced archives, such as SLAAP and OSCAAR, designed specifically around collections of speech recordings, allow more thorough, more flexible, and faster access to large databases of speech recordings than other techniques, such as traditional file management-based organizational schemes. It is hoped that this short exposition has illustrated some of the benefits possible through the development of feature-rich web-based repositories for audio, spoken language data.

Acknowledgments

OSCAAR is being developed in the Speech Communication Research Lab at Northwestern University (Ann Bradlow, Principal Investigator) and is funded by NIDCD grants R01DC005794 and R56DC005794, National Institute of Health, National Institute of Deafness and Other Communication Disorders, and The Hugh Knowles Center for Clinical and Basic Science in Hearing and Its Disorders at Northwestern University. SLAAP has been made possible through support from the North Carolina State University Libraries, the North Carolina Language and Life Project (Walt Wolfram, Director), and the William C. Friday Endowment at NC State University.

References

Bird, Steven and Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79 (3): 557-82.

[13] cf. Kendall 2007.

[14] Kendall 2008, 2009.

Boersma, Paul and David Weenink. 2009. Praat: Doing Phonetics by Computer [Computer Program].

Carter, Phillip. 2009. Speaking subjects: Language, subject formation, and the crisis of identity. Doctoral diss., Duke University.

Kendall, Tyler. 2007. The Sociolinguistic Archive and Analysis Project: Empowering the sociolinguistic archive. *Penn Working Papers in Linguistics* 13(2): 15-26.

Kendall, Tyler. 2008. On the history and future of sociolinguistic data. *Language and Linguistics Compass* 2 (2): 332-351.

Kendall, Tyler. 2009. Speech rate, pause, and linguistic variation: An examination through the Sociolinguistic Archive and Analysis Project. Doctoral diss., Duke University.

Kendall, Tyler and Valerie Fridland. 2010. Mapping production and perception: The influence of regional and individual norms. Paper presented at the Linguistic Society of America 2010 Annual Meeting, in Baltimore, MD.

Kendall, Tyler, Gerard Van Herk, and Joan Bresnan. 2009. The dative alternation in African American English: Researching syntactic variation and change in a conglomerated sociolinguistic corpus. American Association for Corpus Linguistics 2009 Annual Meeting, in Edmonton, Alberta.

Kohn, Mary. 2008. *Latino English in North Carolina: A Comparison of Emerging Communities*. Masters thesis, North Carolina State University.

Mallinson, Christine and Tyler Kendall. 2009. "The Way I Can Speak for Myself": The Social and Linguistic Context of Counseling Interviews with African American Adolescent Girls in Washington, DC. In *African American Women's Language,* ed. S. L. Lanehart, 110-126. Newcastle upon Tyne: Cambridge Scholars Press.

Newman, John. 2008. Spoken corpora: Rationale and application. *Taiwan Journal of Linguistics* 6 (2): 27-58.

R Development Core Team. 2009. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. [Computer Language].

Van Engen, Kristin, Melissa Baese-Berk, Rachel Baker, Arim Choi, Midam Kim, and Ann Bradlow. *In press*. The Wildcat Corpus of Native- and Foreign-Accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and Speech*.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago