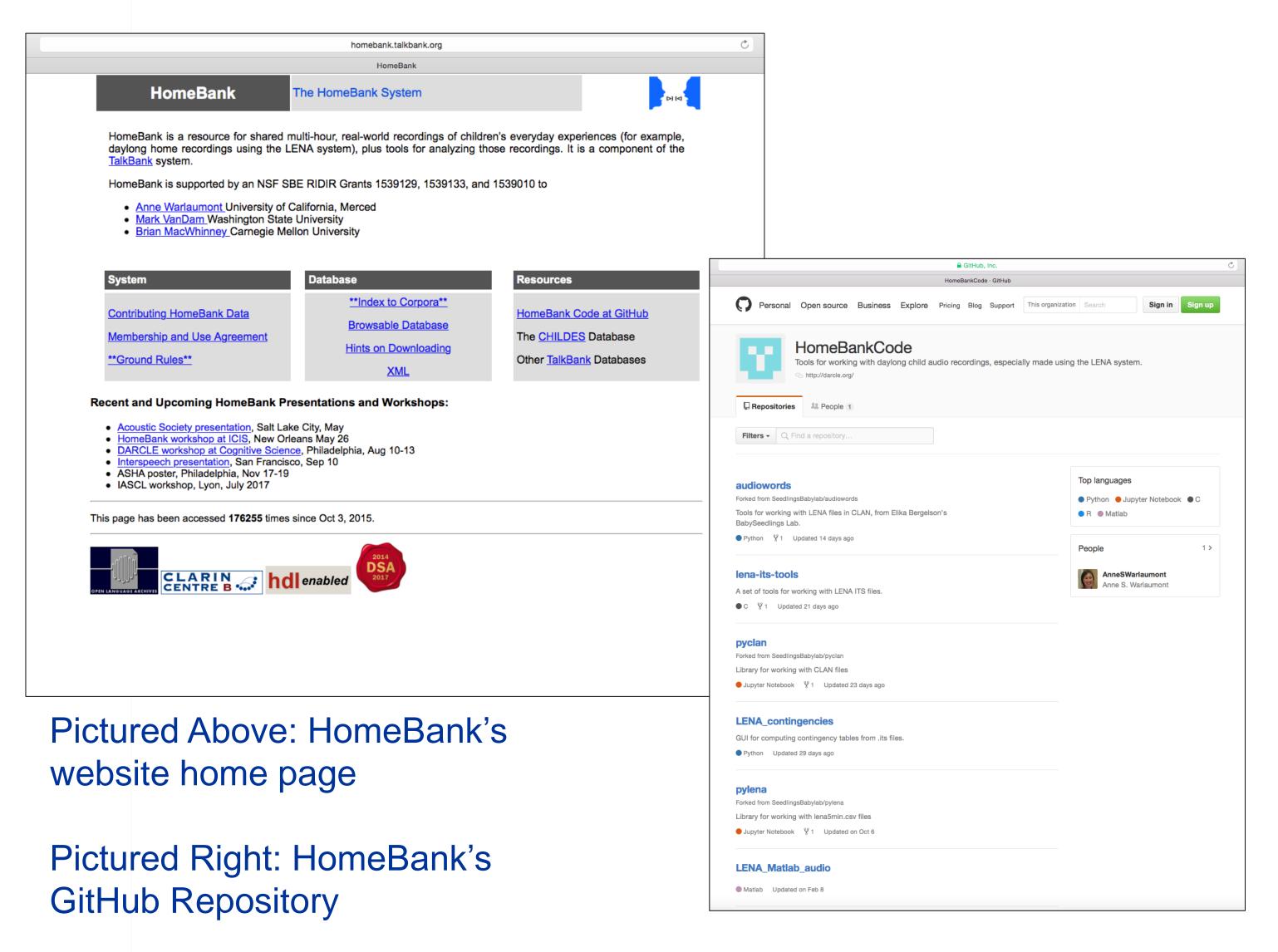
HomeBank: A resource for sharing daylong audio recordings from children and families

Gina M. Pretzer, Anne S. Warlaumont, Mark VanDam, Brian MacWhinney, Christopher J. Reding

What is HomeBank?

- HomeBank (<u>http://www.homebank.talkbank.org</u>) is an NSF-supported project that maintains an online repository which provides a formal system for archiving and sharing daylong audio recordings, and associated transcripts and metadata (see VanDam et al., 2016)
- GitHub repository http://www.github.com/HomeBankCode) hosts computer code for LENA analysis tools written by contributing researchers, as well as documentation for selected scripts.
- 26 members (plus associated RAs and graduate students) and 5 contributors from 4 different countries
- HomeBank is supported by an NSF SBE RIDIR Grants 1539129, 1539133, and 1539010



Interested in gaining access to HomeBank data or contributing data of your own?

Send an email to 4homebank@gmail.com or visit www.homebank.talkbank.org for more information

Online Repository of Daylong Audio Recordings

- HomeBank's online reposit comprised of two levels: P and Private
- The repository currently ho corpora – 1 Public, 7 Privat
- The corpora include audio recordings, related transcri (in CHAT format), and meta (e.g., demographic data, S family status, disorder, etc.
- Built in CLAN window to analyze transcripts

GitHub Repository

- Researchers from several have contributed original co for LENA analysis tools.
- Included are scripts for exc audio segments from LENA recordings, analyzing the L .its files, and anonymizing files written in Python, Matl Perl, and Shell.
- Documentation for selected scripts containing step-by-s instructions and pictures detailing how to use these





| f | Public Repository Anyone can access the data in this repository |
|---------------------------------------|--|
| itory is Public olds 8 | All vetted materials – all sensitive and potentially identifying information has been removed Last names, SSN, address, date of birth, etc. |
| ate ripts tadata SES, c.) | Private Repository Requires membership for access – users must agree to HomeBank's User Agreement Raw data – sensitive or identifying information may remain |

| Researchers can create output files with specific information about the automatically labeled vocalizations, such as frequencies of vocalizations by different speakers, utterance |
|--|
| onset/offset times, and frequencies of child-adult and |
| adult-child exchanges |
| Some scripts use programs such as PRAAT to glean acoustical information (e.g., |
| fundamental frequency, intensity |
| etc.) for vocalizations within recordings |
| |
| |



| | | homebank.talkbank.org | |
|--|---|---|--|
| _ | _ | HomeBank | |
| HomeE | Bank | The HomeBank System | ым |
| data has been diarized bu on how to apply for Home <u>NC-SA 3.0 License.</u> The | ut not transcribe Bank members use of private F | format. Transcribed data have been cleared for public access. The extered. Those data are password protected and open only to HomeBank m ship, see <u>this link</u> . The use of public HomeBank data is governed by the HomeBank data is governed by the HomeBank data is governed by the <u>HomeBank Data Use Agreement for</u> ation that accompanies each corpus and follow the guidelines for <u>data-</u> | embers. For information <u>Creative Commons BY-</u> <u>r Unvetted Recordings</u> . |
| This page provides an inc | dex to HomeBa | nk data. In the Corpus Column there are three links. | |
| The second link will | I download the | escription of the corpus in your browser window. transcripts. laying the downloadable media. | |
| You can also browse the | entire HomeBa | nk database online from <u>this link</u> . | |
| Corpus | Recordings | Description | Contributor |
| Bergelson Seedlings Transcripts Audio | 49 | Study of exposure to object words | Elika Bergelson |
| McDivitt Transcripts Audio | 25 | Children born to adolescent and adult mothers | Karmen McDivitt |
| VanDam Public Transcripts Audio | 53 | 159 5-minute transcribed segments from 53 daylong recordings | Mark VanDam |
| VanDam1 Transcripts Audio | 9 | Preschoolers with and without hearing loss | Mark VanDam |
| VanDam2 Transcripts Audio | 74 | Description not yet ready | Mark VanDam |
| VanDam3 Transcripts Audio | 788 | Description not yet ready | Mark VanDam |
| Warlaumont Transcripts Audio | 18 | Longitudinal study of children learning English and Spanish | Anne Warlaumont |
| Winnipeg Transcripts Audio | 8 | Comparison across different child care settings | Melanie Soderstrom |
| | | | |

 Access to these data is especially helpful for researchers who are unable to access specific populations (e.g., age, clinical diagnoses, etc.) or who do not have the financial or other resources to collect their own recordings

| 1 contribu 35 lines 1 # 2 # 3 # 4 # 5 # 6 # | tzer Updated instructions to include lin utor (33 sloc) 2.07 KB | e numbers | | | caf42de on Apr 1 | |
|--|---|--|---|---|--|-----------------------------------|
| 35 lines 1 # 2 # 3 # 4 # 5 # 6 # | | | | | | |
| 1 # 2 # 3 # 4 # 5 # 6 # | (33 sloc) 2.07 KB | | | | | |
| 2 # 3 # 4 # 5 # 6 # | | | R | aw Blame History | | |
| 3 # 4 # 5 # 6 # | # Gina M. Pretzer | | | | | |
| 5 # 6 # | # # This tool allows you to run the " | readits_start_end_content.pl" tool on multip | ole .its files. | | | |
| 6 # | | its file is written into an individual .csv | file. | | | |
| 7 # | ∓ # Takes the following command line : | arguments: | | | | |
| | | (e.g. "~/lena-its-tools/RunFolder_readits.s | sh") | | | |
| 8 # 9 # | <pre># # Instructions:</pre> | | | | | |
| | | sired .its files in a folder of its own. You | u may nest folders for | each participant wit | hin the overall fo | |
| 11 # | | pants" folder holds subfolders for each part | | * | <u> </u> | |
| | # 2.) In line 25, enter the path to # 3.) In line 26, specify any subfo | the main folder containing the desired .its lders within the main folder | s files (e.g. "~/deskt | op/Gina/Participants" |) | |
| 14 # | | med starting with WW, so the * allows for al | ll the WW folders to b | e opened | | |
| | <pre># 4.) Specify the .its file # 5.) Norma the output files (lines.</pre> | 20. 21) | | | | |
| 16 # 17 # | <pre># 5.) Name the output files (lines # This replaces the ".its"</pre> | 29-31) ending of the filename to rename the file, a | and changes the ".its" | suffix to ".txt", si | nce the output fil | |
| | # 6.) In line 32, set the path for | the "readits_start_end_content.pl" file (e.g | - | | - | |
| | # 7.) Save all changes # 8.) Launch Terminal | | | | | |
| | <pre># 8.) Launch Terminal # 9.) Navigate to directory where "</pre> | RunFolder_readits.sh" is located | | | | |
| | # 10.) Run this file (sh RunFolder_ | readits.sh) | | | | |
| 23 24 | | | | | | |
| 25 c | | n/WWTranscriptionProject/Participants/ ## I | | icipants" is the fold | er that contains t | |
| | | s are named for each participant, all beginn | ning with "WW" | | | |
| 27 28 | <pre>do cd \$folder for itsfile in *.it</pre> | s | | | | |
| 29 | | =`echo \$itsfile sed 's/\.its/CHNStartEndUt | | | | |
| 30 31 | | =`echo \$itsfile sed 's/\.its/ANStartAndEnd =`echo \$itsfile sed 's/\.its/CHN AN Segmer | | | | |
| 32 | perl | | | | | |
| 33 34 | done cd | 38 lines (29 sloc) 2.96 KB | | | Raw Blame | e History 🖵 🖉 |
| | | second intervening between the less. This program returns two o vocalizations and responses, and This program requires that you h A B C 1 chnUttDur chnCryVfxDur chnRT previ 2 0 0.21 1.1 NA 3 0 0.6 NA 4 NA NA 0.6 NA 4 NA 0.6 0 1.1 5 0.6 0 1.1 6 0 0.13 NA 7 1 0 NA 8 0 0.28 NA 9 0 1.67 NA 10 0 0.16 NA 11 0.44 0 1.1 | the second gives the second gives the previously run bus and the second gives the second gives the second gives the second gives the second gives the second gives the second gives the second gives the second gives the second gi | t output file gives in information about "readits_start_end_ F G | nformation about t the adult vocalizat _content.pl" on the | the child tions and responses. |
| | | 12 0 0.66 NA 13 0 1.36 NA | 1.1 1.1 1.1 1.1 | 1.1 | | |
| | | 14 0 0.93 NA 15 0.86 0 NA | 1.1 1.1 1.1 1.1 | 1.1 1.1 | | |
| | | A B C I chnUttDur chnCryVfxDur anRT 2 0 0.21 1.1 | D E F | G | | |
| | | 2 0 0.21 1.1 3 0 0.6 1.1 4 0.6 0 1.1 | | | | |
| | | $\begin{array}{c ccccccccccccccccccccccccccccccccccc$ | | | | |
| | | 7 0 0.16 1.1 8 0.44 0 1.1 | | | | |
| | | 9 0 0.66 1.1 10 0 1.36 1.1 | | | | |
| | | 11 0 0.93 1.1 12 0.86 0 1.1 | | | | |
| | | 13 0 0.55 1.1 14 0 0.71 1.1 | | | | |
| | | 15 1.05 0 1.1 16 0.24 0 1.1 | | | | |
| | | How to use "get_binary_RTs.pl" | on a single .its file | (on a Mac): | | |
| | | Launch Terminal | | | | |
| | | Navigate to the directory where the second se | nere "get_binary_R pe the path to the c | | | |