

README file for The Subglottal Resonances Database.

In addition to the information presented in this README file, see Lulich, et al (2012) for further information about the corpus. The proper way to cite this corpus is to cite Lulich et al (2012).

Lulich, Steven M., John R. Morton, Harish Arsikere, Mitchell S. Sommers, Gary K.-F. Leung, Abeer Alwan (2012) *Subglottal resonances of adult male and female native speakers of American English*. *Journal of the Acoustical Society of America* **132**(4):2592-2602.
[\[link\]](#)

Abstract

Subglottal resonances have received increasing attention in recent studies of speech production, perception, and technology. They affect voice production, divide vowels and consonants into discrete categories, affect vowel perception, and are useful in automatic speech recognition. The Subglottal Resonances Database is a new speech corpus of simultaneous microphone and (subglottal) accelerometer recordings of 25 adult male and 25 adult female speakers of American English (AE), between 22 and 25 years of age. The corpus consists of 35 monosyllables (14 “hVd” and 21 “CVb” words, where C is [b, d, g], and V includes all AE monophthongs and diphthongs) in a phonetically neutral carrier phrase (“I said a ____ again”), with 10 repetitions of each word by each speaker, resulting in 17,500 individual microphone (and accelerometer) waveforms. The target vowel in each utterance has been hand-labeled to indicate the start, stop, and steady-state parts of the vowel. For diphthongs, the steady-state refers to the diphthong nucleus which occurs early in the vowel. Height and age of each speaker is also included. The corpus contains a total of 70,204 files distributed across 100 folders, and the total size of the corpus is approximately 61 GB. The corpus fills a gap in the literature on subglottal acoustics and will be useful for future studies in speech production, perception, and technology. Work supported in part by NSF grant number 0905250.

Files and Naming Conventions

All files included in the corpus are organized into folders, with two folders for each speaker. The names of these folders have the form <speaker number>_<session number>. For instance, folder 9_1 contains the data from the first session of speaker 9, while folder 9_2 contains the data from the second session of speaker 9. For each speaker, the first session includes the CVb words, while the second session includes the hVd words. Speakers 1-8, as well as ten additional speakers, were not included in the final database. In total, there are 100 folders: 25 folders for the first session of males, 25 for the second session of males, 25 for the first session of females, and 25 for the second session of females.

Each folder for session one contains 843 files, including 1 *.mat file, 210 *_acc.wav files, 210 *_mic.wav files, 210 *_mic.bmp files, 210 *_mic.TextGrid files, and 2 additional files called *ACC1.wav and *ACC2.wav (henceforth “*ACC files”).

- **1 *.mat file.** This is a MATLAB file which contains information about the speaker and the recording session.

- **210 *_acc.wav** files. These are accelerometer recordings of 10 repetitions of each of the 21 CVb words.
- **210 *_mic.wav** files. These are microphone recordings of 10 repetitions of each of the 21 CVb words. Note that the microphone and accelerometer files were recorded simultaneously.
- **210 *_mic.bmp** files. These are bitmap images of the target vowel in each of the 210 utterances including the locations of the start, end, and steady-state labels.
- **210 *_mic.TextGrid** files. These are the Praat TextGrid files containing the start, end, and steady-state labels.
- **2 *ACC** files. These are accelerometer recordings of the sustained vowel [a:] for which excellent signal quality was achieved.

Each folder for session two contains 561 files, including 1 *.mat file, 140 *_acc.wav files, 140 *_mic.wav files, 140 *_mic.bmp files, and 140 *_mic.TextGrid files. As in the session one folders, the *.mat file is a MATLAB file containing information about the speaker and the recording session, and the *_acc.wav, *_mic.wav, *_mic.bmp, and *_mic.TextGrid files are the recordings themselves, the images of the labeled target vowels, and the Praat TextGrids containing the labels for 10 repetitions of each of the 14 hVd words.

As an example, consider the file “s9_2_1_2_14-Jan-2010_acc.wav”. The initial “s” simply stands for “speaker”, and the following numbers “9_2” are identical to the folder name - they are the speaker and session numbers. The next number (1) indicates the utterance number, and the next one indicates the repetition number (2). The date on which the recordings were made is then appended (14-Jan-2010), followed by an indication that this is an accelerometer signal file. Note that some repetitions may have been discarded and repeated, so that the repetition number is not always between 1 and 10. For instance, for speaker 9, session 2, utterance 1, there is no 10th repetition (s9_2_1_10_14-Jan-2010_acc.wav), but there is an 11th repetition instead (s9_2_1_11_14-Jan-2010_acc.wav).

For session 1 the utterance numbers correspond to the following CVb words:

Number	Word	Number	Word	Number	Word
1	beeb	8	deeb	15	geeb
2	beb	9	deb	16	geb
3	bob	10	dob	17	gob
4	boob	11	doob	18	goob
5	bibe	12	dibe	19	gibe
6	bowb	13	dowb	20	gowb
7	boib	14	doib	21	goib

For session 2 the utterance numbers correspond to the following hVd words:

Number	Word	Number	Word	Number	Word
1	heed	6	hod	11	hide
2	hid	7	hud	12	how'ed
3	hay'ed	8	hode	13	hoid
4	head	9	hood	14	heard
5	had	10	who'd		

The *.mat Information Files

The information files contain information about the speaker and the recording session under four variables, of which the variable `speakerInfo` is of greatest importance for acoustic analysis of the recordings:

- **displayInfo**

This variable has three fields indicating the properties of the sound recordings as displayed to the experimenters during live recording:

fs: The sampling rate. Typically, $fs = 16,000$.

bitRate: The bit rate. Typically, $bitRate = 16$.

nChannels: The number of channels. This is typically 2.

- **recordingInfo**

This variable has the same structure as `displayInfo`, but presents the sampling rate, bit rate, and number of channels of the sound recordings as actually made and saved to file. Sampling rate is 48,000 Hz, bit rate is 16, and the number of channels is 2.

- **sentenceInfo**

This variable contains information about the utterances produced by the speakers:

wordListFileName: The name and relative location of the text file used to generate the word list to be recorded.

start: The beginning part of the carrier phrase.

end: The end part of the carrier phrase.

wordList: The list of words in the `wordListFileName`. This includes both 'I said a' and 'again' as the first two entries, even though they are separately recorded under the *start* and *end* variables.

presentationID: The ID (utterance number) of each word to be recorded, in random order. For example, in session 1 there are 21 utterances. For speaker s68, the first utterance recorded was "I said a doib again", with word ID 14.

utteranceOK: Either 1 or 0. Utterances deemed acceptable to the experimenter (i.e. no background noise, no slips of the tongue, etc.) were assigned a value of 1. Utterances deemed unacceptable were assigned a value of 0, and were repeated at the end of the experiment. For example, speaker s68 in session 1 had two utterances of "I said a beeb again" rejected. These utterances were appended to the end of the experiment.

repetitionNumber: The indication of how many times a word had been presented to the speaker for recording. Early in each session, the value is usually 1 since most words have been presented only once. Late in each session, the value is usually 11

since most words have been presented the full 11 times. Words which received a value of 0 for *utteranceOK* were appended at the end, and therefore have *repetitionNumber* greater than 11. A minimum of 11 repetitions of each word was chosen in order to guarantee that at least 10 high quality recordings of each utterance would be available for inclusion in the corpus (for instance, in case of experimenter error in assigning values to *utteranceOK*).

- speakerInfo

This variable contains information about the speaker:

ID: The speaker ID.

Session: The session number (either 1 or 2).

Gender: Either 'male' or 'female'

Height: Height in feet and inches (if specified) or centimeters.

Age: Age in months.

nativeLanguage: The native language of the speaker.

testLanguage: The language of the test sentences.