Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2016;68:107–111
DOI: 10.1159/000452128

Published online: October 27, 2016

Arizona Child Acoustic Database Repository

Kate Bunton Brad H. Story

Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, AZ, USA

Keywords
Longitudinal recordings · Speech development · Vowels

Abstract
Objective: The goal of the Arizona Child Acoustic Database project was to obtain a large set of acoustic recordings, primarily vowels, collected from a cohort of children over a critical period of growth and development. ***Method:*** Data was recorded longitudinally from 63 children between the ages of 2;0 and 7;0 at 3-month intervals. The protocol included individual American English vowels and diphthongs, nonsense multi-vowel transitions, word level multi-vowel sequences (e.g., Hawaii), single-syllable words targeting each American English vowel, short sentences, and conversation. ***Results:*** Acoustic files are available for download through the University of Arizona Library Repository for use in future research projects. ***Conclusion:*** Longitudinal recordings may be of interest because they allow tracking of acoustic characteristics produced by an individual child during a period of rapid growth and speech development.

Introduction

The goal of the Arizona Child Acoustic Database (ACAD) project was to obtain a large set of acoustic recordings, primarily vowels, collected from a cohort of children over a critical period of growth and development. Five groups of typically developing children from American English-speaking families were initially recruited to participate. Each group was representative of ages 2, 3, 4, 5, or 6 years. Data were collected longitudinally with children being recorded at 3-month intervals until they reached approximately the age of 7;0. This approach ensured that the database would include cross-sectional as well as longitudinal data. For our purposes, longitudinal recordings were of particular interest because they allow tracking of acoustic characteristics produced by an individual child during a period of rapid growth and speech development.

Collection of this database was part of a larger research project that included the development of a computational model of speech production. The original intent was for the database to serve the needs of the modeling effort, which was primarily focused on vowels. The protocol was largely developed to provide a wide range of vowels embedded in various phonetic contexts. Nonetheless, many samples are included for which analysis of consonants could be performed.

KARGER

E-Mail karger@karger.com
www.karger.com/fpl

© 2016 S. Karger AG, Basel
1021–7762/16/0683–0107$39.50/0

Kate Bunton, PhD
Department of Speech, Language, and Hearing Sciences
University of Arizona, PO Box 210071
Tucson, AZ 85721 (USA)
E-Mail bunton @ email.arizona.edu

Table 1. Age at the initial and final recording sessions for individual speakers

Speaker	Sex	Age at initial recording, years;months	Age at final recording, years;months	Recordings available, n
1	F	5;8	7;0	5
2	F	2;5	5;11	12+
3	F	3;8	4;4	3
4	F	3;7	6;10	10
5	M	3;7	6;10	10
6	M	5;5	6;10	5
7	M	3;9	6;10	10
8	M	2;9	6;4	11+
9	F	4;0	7;0	10
10	M	4;2	6;11	8
11	F	3;5	6;5	10
12	F	6;6	6;11	2
13	F	6;6	6;11	2
14	M	2;0	5;4	10+
15	F	2;7	5;0	8
16	M	2;1	2;1	1
17	M	5;8	6;8	4
18	F	2;8	6;0	9+
19	F	3;8	7;0	10
20	F	6;6	6;10	2
21	M	3;10	5;2	5
22	M	2;1	2;5	3
23	F	3;4	6;6	10+
24	F	3;2	4;10	6
25	F	4;1	4;5	2
26	M	4;11	5;3	2
27	F	2;8	2;8	1
28	M	4;2	7;0	9
29	M	2;0	4;10	9
30	M	no data		0
31	F	3;4	3;4	1
32	M	6;1	6;9	3
33*	M	3;8	6;6	10
34	M	4;6	6;9	8
35	M	4;1	5;5	5
36	F	5;3	6;7	5
37	F	3;3	3;7	2
38	M	2;9	4;6	6
39	M	2;10	4;8	6
40	M	4;9	6;6	6
41	M	2;0	4;4	8
42	F	4;8	4;8	1
43	M	5;7	6;11	5
44*	F	2;6	4;3	6
45	F	4;4	6;5	7
46	M	5;2	6;11	6
47	M	3;5	5;10	8+
48	F	3;2	5;10	9
49	F	2;2	3;4	3
50	M	6;10	6;10	1
51*	F	4;9	6;3	5
52	M	2;0	4:5	8+
53	M	2;0	4;4	7+
54	F	2;2	4;2	7+
55	F	1;11	3;11	6+
56	M	2;3	2;6	2
57	F	2;6	4;3	6+
58	F	2;6	4;3	6+
59	F	2;6	4;4	6+
60	F	2;1	2;6	2
61	M	5;5	6;9	5
62	F	4;6	5;11	4+
63	M	4;3	4;7	2
64	F	2;4	3;0	2+

* The speaker was determined to have a speech sound disorder. + The speaker is continuing to participate in the protocol.

Method

The protocol for this project was approved by the Institutional Review Board at the University of Arizona. Legal guardians provided written consent for the participation and inclusion of their child's data in an online repository. Child assent was waived based on the speaker's age; however, all speakers were told that they could discontinue participation at any time. Speakers were able to choose a small toy at the end of each session and guardians were provided monetary compensation.

Speakers

ACAD contains recordings from 63 speakers (33 females, 30 males) between the ages of 2;0 and 7;0. Speakers were recorded producing the same protocol at 3-month intervals until they reached the age of 7;0. Table 1 reports the sex of each speaker, the age at the initial and final recordings, and the number of recordings currently available. A plus sign next to the number of recordings available indicates that a particular speaker is still active and additional recordings will be added. Note that there is no data available for Speaker 30 because this speaker was used for training purposes and not included in the database; thus, the number "30" serves simply as a placeholder. There were 25 speakers who began recording between the ages 2;0 and 2;11, 14 speakers between 3;3 and 3;10, 12 speakers between 4;0 and 4;9, 7 speakers between 5;2 and 5;7, and 5 speakers between 6;1 and 6;10. This distribution of speakers allows the dataset to be used to address both cross-sectional and longitudinal research questions.

Speakers passed a hearing screening at 20 dB HL for frequencies 0.5, 1, 2, and 4 kHz in both ears [1]. Hearing was rescreened

Table 2. Excerpt of the spreadsheet (based on 4 speakers) available on the ACAD repository site detailing the number of sessions available for each speaker and the exact age at recording (years;months)

Speaker	Gender	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12
ACAD1	F	5;8	6;1	6;4	6;8	7;0							
ACAD2	F	2;5	2;8	3;0	3;4	3;7	4;0	4;4	4;8	4;11	5;3	5;7	5;11
ACAD3	F	3;8	4;0	4;4									
ACAD4	F	3;7	3;11	4;5	4;9	5;2	5;6	5;10	6;2	6;6	6;10		

every 12 months. Speakers (and their guardians) self-reported being a monolingual, native, American English speaker and having no history of speech or language difficulties. During the course of the recording session, the investigator (K.B.) had concerns about the speech production skill of 3 speakers (ACAD33, ACAD44, and ACAD54). All 3 of these speakers participated in a clinical assessment for speech sound disorders with the investigator (K.B., a certified speech-language pathologist) and were referred to their local school district for services. Recordings from these speakers are included in the database but are tagged as having a speech sound disorder in the online database (indicated in the table with an asterisk). Speaker 48 has a distorted /r/ production but does not meet the criteria for diagnosis of a speech sound disorder based on her age.

One highlight of the database is the availability of longitudinal data for a number of speakers. 11 of the speakers in the database have 10 or more recordings (Table 1). This is significant because data for these speakers span at least 30 months, representing a significant period of growth. 10 speakers in the database have data from a period of 24–27 months (8 or 9 recordings). A spreadsheet detailing the recordings available for each speaker can be found on the repository website; this file continues to be updated as more recordings are completed. An example of the spreadsheet has been reproduced as Table 2 based on 4 speakers. The table lists the age of the speaker at each recording session in years and months (y;m). As an example, Speaker 1 has 5 sessions available between the ages 5;8 and 7;0, whereas Speaker 2 has 12 sessions beginning at the age of 2;5 and continuing until 5;11. This speaker is continuing to participate, and additional recordings will be available.

Speech Tasks

The speech tasks included in the protocol are shown in the Appendix. The format of the Appendix lists the record numbers in the database, which correspond to each speech task. The protocol included individual American English vowels and diphthongs /i,ɪ, e,ɛ,æ,ə,ɝ,ʌ,u,ʊ,o,ɔ,ɑ,aɪ,aʊ,ɔɪ/, nonsense multi-vowel transitions (e.g., /iɑui/), word level multi-vowel sequences (e.g., Hawaii), single-syllable words targeting each American English vowel, short sentences, and in some cases, conversation. Conversation was typically initiated by the child and therefore, topics varied widely.

All speakers were prompted to produce the speech targets either using graphical prompts/written words presented on a computer screen, or verbal prompts by the investigators. Prompts were dependent on the child's age and abilities (i.e., younger children could not read). Since data were collected longitudinally, most children used only graphical or text prompts during repeated visits. Speakers were instructed to speak each word clearly, and the pace of presentation was controlled by the investigator who advanced the prompts on the computer. If the investigator judged a word production to be irregular, e.g., using a silly voice, the child was prompted to repeat that word at the end of the protocol.

Recording Procedures

Speakers were seated in a sound-treated room with a microphone (AKG SE 300B) placed 10 cm in front of their mouth. Signals were recorded digitally using a Marantz PMD671, 16 bit PCM (uncompressed) at 44.1 KHz. A calibration tone was recorded at the start of each recording session and is available in the database as record1.wav. This file consists of the tone as well as the investigator's verbal statement of the sound pressure level in dB SPL. All recordings were manually edited using Praat [2] to remove extended silence, pauses, and prompting cues by the investigators. The Marantz recorder was found to generate some low-amplitude noise below 100 Hz. To eliminate it, each signal was postprocessed in Matlab (Mathworks, 2015) with a 512 point FIR high-pass filter with a cutoff frequency located at 100 Hz. The tasks were saved using a standard naming convention so that all files with the same name (e.g., record2.wav) contain the same speech tasks across speakers and recording sessions. If a particular task is missing, an adult voice stating "no data" was inserted into the record in place of the target. The naming conventions for the speech tasks are included in the Appendix.

Database

The ACAD repository is hosted by the University of Arizona library (http://arizona.openrepository.com/arizona/handle/10150/316065) and can be freely accessed following registration. The database contains a set of 31 wav files for each speaker from each recording session. Files from a single recording session are stored as a zip file in the repository to facilitate storage and download. Files are named using the standard convention ACADnumber_sex_age in year_age in months_record number, for example Speaker ACAD3, male, at the age of 5;7 for record 14 would be named as ACAD3_m_5y7m_r14. Files can be searched within the repository by speaker number, age, or gender. All files can be downloaded for further use. As noted in Table 1, the 16 speakers with a plus sign next to their age at the final recording are continuing participation, and additional acoustic files will be uploaded to the database for these speakers. To obtain instructions and permission to access ACAD, email Kate Bunton (bunton@email.arizona.edu).

Discussion

The ACAD was assembled primarily with the goal of facilitating studies of the development of vowel production in children. The protocol provides audio samples appropriate for measuring formant frequencies of children longitudinally as they develop over several years or across speakers at a specific age. These could include typical measurements at the midpoint of syllables as well as time-varying formants over the duration of a syllable or word. In either case, the measurement of formant frequencies based on the same utterances produced by the same child over several years is fairly unique. Typically, vowel development studies are carried out with a cross-section of data collected from many children of different ages. In addition, the number of samples collected for each speaker should be sufficient to observe the range of fundamental frequency used at each age increment.

The nonmeaningful vowel transitions and repetitions, such /iɑui/ and /iɑiɑ/, respectively, were included in the protocol as a means of eliciting a smoothly changing, unoccluded vocal tract shape, resulting in continuous time-varying formant frequencies expressed in the recorded acoustic signal. The temporal characteristics of the formant frequencies, such as transition slope and curvature, as well as the (F1, F2) vowel space trajectory shape, may provide some insight into the motor control abilities of a speaker. When measured for the same speaker several times over the course of speech development, these measurements could be compared to determine their characteristics across age. Additionally, since time-varying formant frequencies have been shown to correspond to a coordinated pattern of vocal tract shaping patterns [3, 4] for adults, it would be of interest to perform similar mappings for a child-like vocal tract model [5] to determine if the coordination of the shaping patterns is similar or different from that in adults.

The vowel-consonant-vowel sequences and words provide cases for which the temporal variation of the formants is due to the influence of both vowels and consonants on the time-dependent vocal tract configuration. Measurements of formant transitions over the duration of a vowel-consonant-vowel sequence or word, particularly, at the onset and offset of consonants, could be used to investigate aspects of coarticulation for specific speakers as they develop speech production abilities. These data could also facilitate the development of computational models of a child-like vocal tract [5, 6].

It is noted that since the protocol was primarily comprised of word lists, the database does not lend itself to analyses of speech rate or articulation rate. Some speakers produced a fair amount of spontaneous speech from which some estimates of rate could be measured, but it is not likely that there is enough material at each age increment to allow for a thorough study.

Acknowledgements

This paper was supported by NSF BSC-1145011. The authors thank Megan Kittleson and Kim Neely for their help in building the online repository.

Appendix

ACAD task list showing standard convention for naming acoustic records
Record 1: calibration tone, investigator's verbal statement of dB SPL
Record 2: green red black yellow blue white pink purple orange brown
Record 3: feet sheep bee beep
Record 4: sustained /i/, iu ibu idu igu ilu iru iɑ ibɑ idɑ igɑ ilɑ irɑ
Record 5: face cake baby table
Record 6: sVd syllables: sid sɪd sed sɛd sæd sɑd sɔd sʊd sud sod sʌd
Record 7: bed head pet wet
Record 8: pig kiss fish chicken
Record 9: cat sad hat dad
Record 10: hVd syllables: hid hɪd hed hɛd hæd hɑd hɔd hʊd hud hod hʌd
Record 11: sock hop on pop top fox
Record 12: sustained/ɑ/, ɑu ɑbu ɑdu ɑgu ɑlu ɑru ɑi ɑbi ɑdi ɑgi ɑli ɑri
Record 13: dog talk walk hawk
Record 14: duck bug cup tub
Record 15: soup boot suit food
Record 16: sustained/u/, uɑ ubɑ udɑ ugɑ ulɑ urɑ ui ubi udi ugi uli uri
Record 17: boat goat coat go
Record 18: one two three four five six seven eight nine ten
Record 19: book foot good hook
Record 20: vowel transitions: iɑui iuɑi uiɑu oæiu
Record 21: vowel repetitions: iɑiɑiɑiɑ ɑiɑiɑiɑi oaeoaeoae aeoaeoaeo iuiuiu uiuiuiu
Record 22: Arizona Iowa Ohio Hawaii
Record 23: cow house owie
Record 24: bike kite five
Record 25: bird purple dirt earth worm burp
Record 26: boy toy voice
Record 27: You were away
Record 28: To feed the cat one must shoo the dog
Record 29: Buy bobby a puppy
Record 30: abracadabra
Record 31: the end
Record 32: miscellaneous single words
Record 33: spontaneous speech
Record 33b, 33c, 33d, 33e more spontaneous speech as needed

If a target utterance was not elicited from the speaker, the correctly named file will still exist, but the recording will include an adult voice saying "no data."

References

1 American Speech-Language-Hearing Association: Guidelines for identification audiometry. ASHA 1985;27:49–52.

2 Boersma P, Weenink D: Praat: doing phonetics by computer. 2013, version 5.3.51. http://www.praat.org/ (retrieved June 2, 2013).

3 Story BH: Time-dependence of vocal tract modes during production of vowels and vowel sequences. J Acoust Soc Am 2007;121: 3770–3789.

4 Story BH, Titze IR: Parameterization of vocal tract area functions by empirical orthogonal modes. J Phonetics 1998;26:223–260.

5 Story BH, Bunton K: Formant measurement in children's speech based on spectral filtering. Speech Commun 2015;76:93–111

6 Story BH: Phrase-level speech simulation with an airway modulation model of speech production. Comput Speech Lang 2013;27: 989–1010.