

Introduction

The American English Map Task (AEMT) database is a collection of 16 dialogues recorded using the Map Task protocol in July 1999. With this protocol, data were recorded from two subjects as they worked together to complete the task of navigating a map. (For more information about the map task, please see <http://www.hcrc.ed.ac.uk/maptask/>.) The data were recorded to DAT in a sound-attenuated chamber at the MIT Speech Communication Group. The recordings were made over the course of two sessions, and at the end of each session, the subjects read a list of the landmarks (listed at the end of this document) that appeared in the maps, yielding an additional eight recordings.

The data were transferred from the DAT to computer files, and were downsampled to 16,000 samples/s. The two channels (one for each subject) of the dialogue files were separated into single (mono) files, yielding 32 files. The maps that the subjects used, including their markings and annotations have been scanned into .pdf files and are included in this database. Thus, the database is comprised of 40 audio files and 32 .pdf files.

The subjects for the recordings were eight young adult females. All were students in the Boston area, but they grew up in various parts of the country.

Experiment Design

In designing the experiment, the procedures described in McAllister et al. (1990) were followed.

The Maps. Out of the eight sets of maps created for the HCRC Map Task Corpus (Anderson et al., 1991), the quartet 1 (Qtr1) set of maps was used in the AEMT experiment. The four phonological modifications (reduction types) that these maps were designed to elicit were coded as follows:

- 1 -- t-deletion
- 2 -- glottalization
- 3 -- d-deletion
- 4 -- nasal assimilation

Each of these reduction types appears on the maps as a pair of landmarks in four contrast/match conditions (see McAllister et al. (1990) for more detailed explanation of reduction-type coding conventions). Naturally, these data can be used to study phenomena other than the four modifications that the maps were intended to elicit.

Eye Contact. During the recording sessions the subjects were separated from each other with an opaque cardboard screen, thus the experiments were conducted under the no eye contact condition.

Speaker Familiarity. The subjects knew each other before they participated in the experiment. Therefore, the familiarity condition was not used.

Reading. After the experiment, the subjects were asked to read a list of landmark names including those used on the map.

Subjects. The subjects were grouped into two quadruples or "quads". Every participant was exposed to three different maps, one as an instruction giver (IG) and two as an instruction follower (IF). (Note that each subject used the same map for her two turns as an instruction giver, but used two different maps for her turns as an instruction follower.) The subjects were assigned to each dialogue according to the scheme described in McAllister et al. (1990), where **a1**, **a2** refers to one pair of speakers, and **b1**, **b2** refers to the second pair of speakers in a quad. Here is the layout of the two recordings conducted:

Quad1 (Recorded 19-Jul-99)

Map	Contrast	Dialogue	Giver	Follower	Reduction Type
1	"+/+"	1	s1 (a1)	s2 (b1)	++1 (t-deletion)
2	"+/-"	2	s3 (b2)	s4 (a2)	+ -2 (glottalization)
3	"-/+"	3	s4 (a2)	s1 (a1)	- +3 (d-deletion)
4	"-/-"	4	s2 (b1)	s3 (b2)	--4 (nasal assimilation)
3	"-/+"	5	s4 (a2)	s3 (b2)	- +3 (d-deletion)
4	"-/-"	6	s2 (b1)	s1 (a1)	--4 (nasal assimilation)
1	"+/+"	7	s1 (a1)	s4 (a2)	++1 (t-deletion)
2	"+/-"	8	s3 (b2)	s2 (b1)	+ -2 (glottalization)

Quad2 (Recorded 22-Jul-99)

Map	Contrast	Dialogue	Giver	Follower	Reduction Type
1	"+/+"	9	s5 (a1)	s6 (b1)	++1 (t-deletion)
2	"+/-"	10	s7 (b2)	s8 (a2)	+ -2 (glottalization)
3	"-/+"	11	s8 (a2)	s5 (a1)	- +3 (d-deletion)
4	"-/-"	12	s6 (b1)	s7 (b2)	--4 (nasal assimilation)
3	"-/+"	13	s8 (a2)	s7 (b2)	- +3 (d-deletion)
4	"-/-"	14	s6 (b1)	s5 (a1)	--4 (nasal assimilation)
1	"+/+"	15	s5 (a1)	s8 (a2)	++1 (t-deletion)
2	"+/-"	16	s7 (b2)	s6 (b1)	+ -2 (glottalization)

File name convention

The files are named using the following convention:

AEMTdialogue<dial#>_map<map#>_<role>_<subject code>_ms.wav
 AEMTlist_<subject code>_ms.wav
 AEMTdialogue<dial#>_map<map#>_<role>_<subject code>.pdf

where

- subject codes are s1-s8
- dial# is the dialogue number, 1-16

- role is either "if" (information follower) or "ig" information giver
- .wav files contain the audio signals for either the dialogues or the list readings
- .pdf files contain the maps that accompany the dialogue files

(The "ms" in the file names indicates that the file format is Microsoft .wav rather than Klatt .wav, which is commonly used in the Speech Communication Group.)

Because of the size of the data files, we have grouped them in a way that we hope will satisfy those who want to download all of the files, and those who want to pick and choose:

- The read lists are grouped together in a .tar file called AEMTlists.tar
- The dialogue and map files are grouped into tar files by dialogue. For example, the file AEMTdialogue09_map2.tar contains four files, i.e., the audio and .pdf files for both the information giver and the information follower who performed the task during Dialogue 9, using Map 2.

Thus, there are a total of 24 .tar files containing all of the data.

Acknowledgments

The data were recorded by Olga Goubanova, then a graduate student at the Dept. of Linguistics, University of Edinburgh. We are grateful to her for making these recordings and for her documentation of the procedure, on which this README is based.

References

- McAllister, J., Sotillo, C., Bard E.G., and Anderson, A.H. (1990). "Using the map task to investigate variability in speech," *Occasional paper*, Department of Linguistics, University of Edinburgh.
- Anderson, A.H., Bader, M., Bard E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., and Weinhart, R. (1991). "The HCRC Map Task Corpus," *Language and Speech*, v. 34, pp. 351-366.

List of landmarks read by the subjects

1. Abandoned Cottage
2. Antelope
3. Apache Camp
4. Bakery
5. Buffalo
6. Canadian Paradise
7. Canal

8. Canoes
9. Carved Wooden Pole
10. Cattle Ranch

11. Cave
12. Cliffs
13. Collapsed Shelter
14. Crane Bay
15. Desert
16. Diamond Mine
17. East Lake
18. Elephants
19. Farmed Land
20. Fast Flowing River

21. Fenced Meadow
22. Flat Rocks
23. Forge
24. Forrest
25. Forrest Stream
26. Fort
27. Ghost Town
28. Golf Course
29. Graveyard
30. Green Bay

31. Hot Springs
32. Hot Wells
33. Lake
34. Mill Wheel
35. Monument
36. Nuclear Test Site
37. Old Light House
38. Old Mill
39. Old Pine
40. Peak Marker

41. Picket Fence
42. Pine Forrest
43. Rocket Warehouse
44. Roman Baths
45. Round Rocks
46. Saloon Bar
47. Sandstone Cliffs
48. Savannah
49. Site of Plane Crash

50. Slate Mountain
51. Stone Creek
52. Stony Desert
53. Walled City
54. Waterhole
55. West Lake
56. Wheat Fields
57. White Mountain