The ISOLET Spoken Letter Database[†]

Ron Cole
Yeshwant Muthusamy
Mark Fanty

November 29, 1994

1 Description

ISOLET is a database of letters of the English alphabet spoken in isolation. The database consists of 7800 spoken letters, 2 productions of each letter by 150 speakers. Each speaker is identified by a string specifying their gender and initials followed by a number for uniqueness, e.g. "fbjt0" is a female with the initials "bjt." The utterances for each speaker are in a separate directory, one utterance per file.

The speakers are organized into five subsets: ISOLET-1, ISOLET-2, ISOLET-3, ISOLET-4 and ISOLET-5. Each subset contains utterances produced by 30 speakers, 15 male and 15 female. The grouping is arbitrary and roughly chronological. The total space used is 150 megabytes.

2 File Format

The digitized speech files use a format similar to TIMIT [3, 2] adc files. Each file consists of a header followed by a series of 16 bit integers. The header and data are stored in big-endian format with respect to bytes (Sun format); the least significant byte is in the lowest address.

The header has the following format:

No. bytes	Description
2	Size of header in 2 byte words
2	Version
2	Number of channels
2	Rate in quarter micro seconds
4	Number of samples
4	little-endian flag

For ISOLET, the header size is 8 words. The version number is 1. The number of channels is 1. The rate is 250 quarter microseconds per sample, which is 62.5 microseconds per sample, or 16000 samples per second. The little-endian flag is 0.

[†]This research was supported by a grant from Adaptive Solutions Inc., Beaverton, OR and DARPA grant MDDA972-88-J-1004 awarded to the Department of Computer Science, Oregon Graduate Institute. The authors wish to thank Vincent Weatherill for recruiting and recording most of the speakers.

3 Recording Conditions

Speech was recorded in the OGI speech recognition laboratory. The room is 15' by 15' with a tile floor and standard office wall board and drop ceiling. There are two Sun workstations in the room, and three disk drives. The recording equipment was selected to mimic the equipment used to collect the TIMIT database as closely as possible. The speech was recorded with a Sennheiser HMD 224 noise-cancelling microphone, lowpass filtered at 7.6 kHz. Data capture is performed using the AT&T DSP32 board installed in a Sun 4/110. The data is sampled at 16 kHz.

The subjects were seated in front of a Sun workstation and prompted with letters in random order. After each prompt, the subject would strike the *return* key and say the letter. Two seconds of speech were recorded and immediately played back for verification. If the subject spoke too soon or too late and missed the two second buffer, or if the experimenter or subject decided the letter was mis-spoken, the recording would be repeated. There was no attempt to elicit ideal speech. A letter was judged mis-spoken only if there was a significant departure from normal pronunciation.

4 Signal/Noise ratio

We estimated the signal to noise ratio using the following procedure. The digitized waveform was first adjusted by subtracting the mean signal value from each sample so that the new mean is 0. Then the mean amplitude squared is calculated for the center 1/2 of the sonorant (i.e. 1/4 of the sonorant is removed from the beginning and end) and the center 1/2 of the preceding silence. By removing the beginning and end of each segment, we hoped to minimize the transitions. We used the first T utterance from each speaker; the silence before the /t/ burst is usually clean—no breath noise or pre-voicing—so it should reflect the relative background noise well.

$$\sigma_s^2 = \frac{\sum_{i=1}^{i=N} t_i^2}{N}, \text{ for N sonorant samples}$$

$$\sigma_n^2 = \frac{\sum_{i=1}^{i=M} t_i^2}{M}, \text{ for M silence samples}$$

$$\mathrm{S/N} = 10 \log_{10} \left(\frac{\sigma_s^2}{\sigma_n^2} \right)$$

The mean was 31.5 dB, with a standard deviation of 5.6 dB.

5 Signal chopping

In order to save disk space, the silence was removed from each utterance according to the following procedure. The signal was scanned from the ends until "speech" was encountered, the scan then backed out to "silence" and the signal was chopped. The definition of "speech" for the inward scan was 30 consecutive milliseconds of relatively high amplitude or zero-crossing rate. The definition of “silence” for the outward scan was 30 consecutive milliseconds of low amplitude and zero crossing. An additional 50 milliseconds was kept past the beginning and end chop points.

6 Verification

After the recording session, each utterance was verified in two ways. First, the digitized waveform was examined visually to determine if some portion of the utterance was incorrectly deleted by the chop program. If a significant portion of the utterance was deleted, such as the [ch] in “H” or the [ks] in “X,” the utterance was tagged. It was then recovered and chopped by hand. If a small amount of prevoicing or post-vowel voicing was removed, it was not tagged. Second, each utterance was listened to. The listener noted ambiguous utterances and utterances that were incorrectly chopped.

All utterances that were judged as abnormal by the listener were listened to by at least two additional persons. Only utterances that were misperceived by a majority were deleted. Since ISOLET contains only complete speakers (52 letters), removing a letter also meant removing the speaker. In all, 50 letters from 32 speakers were removed from the database.

7 Training and test sets

In our lab, we designated ISOLET1–4 as the training set and ISOLET5 as the test set. Our best results to date [1] are 95% recognition accuracy on ISOLET5 with a network trained on the first token of each letter in ISOLET1–4. For multi-speaker, we used the same net trained on the first token in ISOLET1–4 and tested on the second token for the same speakers. Our recognition accuracy was 96%. During system development, we trained on smaller subsets of ISOLET1–4 (e.g. ISOLET3–4) and tested on the remainder (e.g. ISOLET1–2). This way we avoided unfair tuning of our parameters to the test set ISOLET5.

8 Availability

The ISOLET database can be obtained from the Oregon Graduate Institute by sending a copy of the order form which appears at the end of this report. A small fee is required to cover our costs. The database may be freely copied and distributed.

9 Speaker Information

Subjects were obtained through advertising. Each subject was given a free dessert at a local restaurant in exchange for his or her participation. All speakers reported English as their native language. The ages varied from 14 to 72 years, with an average of 35. A complete listing of each speaker’s age and the state or country where they spent most of their youth (as entered by the subjects) is appended.

References

[1] R. Cole, Mark Fanty, Yeshwant Muthusamy, and Murali Gopalakrishnan. Speaker-independent recognition of spoken english letters. In *International Joint Conference on Neural Networks*, 1990.

[2] W. Fisher, G. Doddington, and K. Goudie-Marshall. The darpa speech recognition research database: Specification and status. In *Proceedings of the DARPA Speech Recognition Workshop*, pages 93–100, 1986.

[3] L. Lamel, R. Kassel, and S. Seneff. Speech database development: Design and analysis of the acoustic-phonetic corpus. In *Proceedings of the DARPA Speech Recognition Workshop*, pages 100–110, 1986.

ISOLET-1

ID	Age	State
fcmc0	38	Oregon
fcmg0	29	Montana
fdcf0	37	Oregon
fec0	35	New Jersey
fet0	60	Florida
fews0	44	New Jersey
fjw0	38	Oregon
fka0	25	Oregon
fkh0	42	Montana
fmb0	31	Michigan
fmbd0	46	USA
fme0	33	Michigan
frw0	22	California
fsaj0	14	Alaska
fskes0	32	California
mjc1	17	Oregon
mjfv0	54	Oregon
mjp0	41	New York
mjrs0	60	New Jersey
mnjh0	26	New York
mnre0	39	Oregon
mrmh1	51	Oregon
mrs0	48	Oregon
msa0	36	Oregon
mtdw0	38	Nebraska
mteb0	32	Washington
mtgr0	39	Florida
mtkm0	41	Iowa
mwmh0	34	Oregon
mwr0	58	Oregon

ISOLET-2

ID	Age	State
facp0	27	Arizona
fbja0	16	Oregon
fbl0	35	Alabama
fdh0	54	Texas
fdlm0	26	Wisconsin
fgw0	47	Tennessee
fhi0	27	Colorado
fjr0	24	Massachusetts
fkma0	24	Oregon
fls0	34	Texas
fmev0	18	Oregon
fplt0	25	USA
frem0	42	California
fss0	26	Oregon
ftmp0	28	Idaho
malb0	28	USA
mdgn0	38	Oregon
mdls0	46	Ohio
mdwh0	42	Utah
mjag0	36	Massachusetts
mji0	26	Hawaii
mjjs0	33	Minnesota
mjs0	25	New York
mjw0	20	Oregon
mjws0	30	California
mls0	47	California
mmaj0	35	Oregon
mrlj0	39	Washington
msdd0	40	Virginia
mtkl0	35	New Jersey

ISOLET-3

ID	Age	State
fah0	43	Washington
famd0	29	California
faw0	29	Illinois
fbj0	18	Oregon
fbjd0	45	Oregon
fbk0	57	Oregon
fcap0	37	New York
fch0	29	California
fch1	45	Utah
fcm0	27	Belgium
fjmr0	22	Oregon
flcb0	33	Oregon
fmlj0	60	Oregon
fms0	29	Japan
fnc0	42	Washington
macj0	42	Montana
mamo0	19	Oregon
mbp0	28	Washington
mdcd0	33	Oregon
mdcd1	51	Missouri
mdht0	35	Missouri
mdlw0	31	Oregon
mdmp0	33	Arizona
mjc0	26	California
mjho0	48	Washington
mmwp0	27	California
mrm0	34	California
mrme0	36	North Dakota
mrmh0	31	Oregon
mrr0	46	Minnesota

ISOLET-4

ID	Age	State
fcb0	31	Oregon
fcc0	42	California
fdle0	31	Michigan
fgh0	26	Hawaii
fit0	25	California
fjar0	53	Washington
fjbc0	44	USA
fjkh0	36	Oregon
fkj0	40	Oregon
fkp0	42	Oregon
fll0	29	New York
fmdf0	20	Oregon
fpe0	39	Missouri
fss1	39	New York
fvca0	36	New York
mbes0	28	Iowa
mce0	29	Alaska
mdhc0	19	Pennsylvania
mdjs0	38	Washington
mgs0	58	Pennsylvania
mhhw0	48	Oregon
mjjh0	17	Oregon
mmgw0	32	Connecticut
mmk0	33	Illinois
mmps0	22	Iowa
mphh0	44	Oregon
mps0	25	California
mrl0	32	California
mtlr0	33	Ohio
mwcs0	44	New York

ISOLET-5

ID	Age	State
farw0	19	New York
fbls0	46	Oklahoma
fceb0	51	Oregon
fel0	41	Oregon
fgs0	34	California
fkf0	41	New Jersey
fkw1	34	Oregon
fla0	44	California
flc0	25	New York
flm0	30	Colorado
fmf0	36	Michigan
fmm1	39	Oregon
fmr0	35	New York
ftlj0	26	Oregon
ftw1	30	New York
mac0	48	New York
mbf0	26	Oregon
mbv0	36	New Hampshire
mcap0	33	Oregon
mcem0	50	Califronia
mcs0	19	Oregon
mjbg0	34	Washington, D.C.
mjgh0	31	Oregon
mpmb0	38	Washington
mrab0	72	Washington
mrs1	33	Oregon
msed0	35	Oregon
mtes0	45	Illinois
mvcw0	39	Oregon
mwjl0	36	California

ISOLET Database Order Form

Medium	Cost	Check to Order
Sun cartridge	$165	
Exabyte 8mm	$50	
1/2in 9 track, 6250 bpi, 2400 ft	$100	
1/2in 9 track, 1600 bpi, 2400 ft	$240	
DEC TK50 cartridge	$200	

Send the order to

Vince Weatherill
Computer Science Dept.
Oregon Graduate Institute
19600 N. W. Von Neumann Dr.
Beaverton Oregon, 97006-1999

Include a check for the appropriate amount payable to Oregon Graduate Institute.

Shipping Address:______________________________
.______________________________________
.______________________________________
.______________________________________
.______________________________________