File: readme.doc, updated 10/12/90

The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus

(TIMIT)

Training and Test Data

NIST Speech Disc CD1-1.1

The TIMIT corpus of read speech has been designed to provide speech data for the acquisition of acoustic-phonetic knowledge and for the development and evaluation of automatic speech recognition systems. TIMIT has resulted from the joint efforts of several sites under sponsorship from the Defense Advanced Research Projects Agency - Information Science and Technology Office (DARPA-ISTO). Text corpus design was a joint effort among the Massachusetts Institute of Technology (MIT), Stanford Research Institute (SRI), and Texas Instruments (TI). The speech was recorded at TI, transcribed at MIT, and has been maintained, verified, and prepared for CD-ROM production by the National

Institute of Standards and Technology (NIST). This file contains a brief

description of the TIMIT Speech Corpus. Additional information including the referenced material and some relevant reprints of articles may be found in the printed documentation which is also available from NTIS (NTIS# PB91-100354).

1. Corpus Speaker Distribution

-- ---------------------------

TIMIT contains a total of **6300** sentences, **10** sentences spoken by each of **630** speakers from **8** major dialect regions of the United States. Table 1 shows the number of speakers for the 8 dialect regions, broken down by sex. The percentages are given in parentheses. A speaker's dialect region is the geographical area of the U.S. where they lived during their childhood years.

The geographical areas correspond with recognized dialect regions in U.S. (Language Files, Ohio State University Linguistics Dept., 1982), with the exception of the Western region (dr7) in which dialect boundaries are not known with any confidence and dialect region 8 where the speakers moved around a lot during their childhood.

**Table 1:** Dialect distribution of speakers

Dialect

Region(dr) #Male #Female Total

---------- --------- --------- ----------

1 31 (63%) 18 (27%) 49 (8%)

2 71 (70%) 31 (30%) 102 (16%)

3 79 (67%) 23 (23%) 102 (16%)

4 69 (69%) 31 (31%) 100 (16%)

5 62 (63%) 36 (37%) 98 (16%)

6 30 (65%) 16 (35%) 46 (7%)

7 74 (74%) 26 (26%) 100 (16%)

8 22 (67%) 11 (33%) 33 (5%)

------ --------- --------- ----------

8 438 (70%) 192 (30%) 630 (100%)

The dialect regions are:

dr1: New England

dr2: Northern

dr3: North Midland

dr4: South Midland

dr5: Southern

dr6: New York City

dr7: Western

dr8: Army Brat (moved around)

2. Corpus Text Material

-- --------------------

The text material in the TIMIT prompts (found in the file "prompts.doc")

consists of 2 dialect "shibboleth" sentences designed at SRI, 450

phonetically-compact sentences designed at MIT, and 1890 phonetically-diverse sentences selected at TI. The dialect sentences (the SA sentences) were meant to expose the dialectal variants of the speakers and were read by all 630 speakers. The phonetically-compact sentences were designed to provide a good coverage of pairs of phones, with extra occurrences of phonetic contexts thought to be either difficult or of particular interest. Each speaker read 5 of these sentences (the SX sentences) and each text was spoken by 7 different speakers. The phonetically-diverse sentences (the SI sentences) were selected from existing text sources - the Brown Corpus (Kuchera and Francis, 1967) and

the Playwrights Dialog (Hultzen, et al., 1964) - so as to add diversity in sentence types and phonetic contexts. The selection criteria maximized the variety of allophonic contexts found in the texts. Each speaker read 3 of these sentences, with each sentence being read only by a single speaker.

Table 2 summarizes the speech material in TIMIT.

**Table 2:** TIMIT speech material

Sentence Type #Sentences #Speakers Total #Sentences/Speaker

------------- ---------- --------- ----- ------------------

Dialect (SA) 2 630 1260 2

Compact (SX) 450 7 3150 5

Diverse (SI) 1890 1 1890 3

------------- ---------- --------- ----- ----------------

Total 2342 6300 10

3. Suggested Training/Test Subdivision

-- -----------------------------------

The speech material has been subdivided into portions for training and

testing. The criteria for the subdivision is described in the file

"testset.doc". THIS SUBDIVISION HAS NO RELATION TO THE DATA DISTRIBUTED ON THE PROTOTYPE VERSION OF THE CDROM.

Core Test Set:

The test data has a core portion containing 24 speakers, 2 male and 1 female from each dialect region. The core test speakers are shown in Table 3. Each speaker read a different set of SX sentences. Thus the core test material contains 192 sentences, 5 SX and 3 SI for each speaker, each having a distinct text prompt.

**Table 3:** The core test set of 24 speakers

Dialect Male Female

------- ------ ------

1 DAB0, WBT0 ELC0

2 TAS1, WEW0 PAS0

3 JMP0, LNT0 PKT0

4 LLL0, TLS0 JLM0

5 BPM0, KLT0 NLP0

6 CMJ0, JDH0 MGD0

7 GRT0, NJM0 DHC0

8 JLN0, PAM0 MLD0

Complete Test Set:

A more extensive test set was obtained by including the sentences from all speakers that read any of the SX texts included in the core test set. In doing so, no sentence text appears in both the training and test sets. This complete test set contains a total of 168 speakers and 1344 utterances, accounting for about 27% of the total speech material. The resulting dialect distribution of the 168 speaker test set is given in Table 4. The complete test material contains 624 distinct texts.

**Table 4:** Dialect distribution for complete test set

Dialect #Male #Female Total

------- ----- ------- -----

1 7 4 11

2 18 8 26

3 23 3 26

4 16 16 32

5 17 11 28

6 8 3 11

7 15 8 23

8 8 3 11

----- ----- ------- ------

Total 112 56 168

4. CDROM TIMIT Directory and File Structure

-- ----------------------------------------

The speech and associated data is organized on the CD-ROM according to the

following hierarchy:

/<CORPUS>/<USAGE>/<DIALECT>/<SEX><SPEAKER\_ID>/<SENTENCE\_ID>.<FILE\_TYPE>

where,

CORPUS :== timit

USAGE :== train | test

DIALECT :== dr1 | dr2 | dr3 | dr4 | dr5 | dr6 | dr7 | dr8

(see Table 1 for dialect code description)

SEX :== m | f

SPEAKER\_ID :== <INITIALS><DIGIT>

where,

INITIALS :== speaker initials, 3 letters

DIGIT :== number 0-9 to differentiate speakers with identical

initials

SENTENCE\_ID :== <TEXT\_TYPE><SENTENCE\_NUMBER>

where,

TEXT\_TYPE :== sa | si | sx

(see Section 2 for sentence text type description)

SENTENCE\_NUMBER :== 1 ... 2342

FILE\_TYPE :== wav | txt | wrd | phn

(see Table 5 for file type description)

Examples:

/timit/train/dr1/fcjf0/sa1.wav

(TIMIT corpus, training set, dialect region 1, female speaker,

speaker-ID "cjf0", sentence text "sa1", speech waveform file)

/timit/test/df5/mbpm0/sx407.phn

(TIMIT corpus, test set, dialect region 5, male speaker, speaker-ID

"bpm0", sentence text "sx407", phonetic transcription file)

Online documentation and tables are located in the directory "timit/doc".

A brief description of each file in this directory can be found in Section 6.

5. File Types

-- ----------

The TIMIT corpus includes several files associated with each utterance. In

addition to a speech waveform file (.wav), three associated transcription

files (.txt, .wrd, .phn) exist. These associated files have the form:

<BEGIN\_SAMPLE> <END\_SAMPLE> <TEXT><new-line>

.

.

.

<BEGIN\_SAMPLE> <END\_SAMPLE> <TEXT><new-line>

where,

BEGIN\_SAMPLE :== The beginning integer sample number for the

segment (Note: The first BEGIN\_SAMPLE of each

file is always 0)

END\_SAMPLE :== The ending integer sample number for the segment

(Note: Because of the transcription method used,

the last END\_SAMPLE in each transcription file

may be less than the actual last sample in the

corresponding .wav file)

TEXT :== <ORTHOGRAPHY> | <WORD\_LABEL> | <PHONETIC\_LABEL>

where,

ORTHOGRAPHY :== Complete orthographic text transcription

WORD\_LABEL :== Single word from the orthography

PHONETIC\_LABEL :== Single phonetic transcription code

(See "phoncode.doc" for description

of codes)

Table 5: Utterance-associated file types

File Type Description

--------- ------------------------------------------------------

.wav - SPHERE-headered speech waveform file. (See the "/sphere"

directory for speech file manipulation utilities.)

.txt - Associated orthographic transcription of the words the

person said. (Usually this is the same as the prompt, but

in a few cases the orthography and prompt disagree.)

.wrd - Time-aligned word transcription. The word boundaries

were aligned with the phonetic segments using a dynamic

string alignment program (see the printed documentation

section "Notes on the Word Alignments" and the lexical

pronunciations given in "timitdic.txt".)

.phn - Time-aligned phonetic transcription. (See the reprint

of the article by Seneff and Zue (1988), in the printed

documentation, and the section "Notes on Checking the

Phonetic Transcriptions" for more details on the phonetic

transcription protocols.)

Example transcriptions from the utterance in "/timit/test/dr5/fnlp0/sa1.wav"

Orthography (.txt):

0 61748 She had your dark suit in greasy wash water all year.

Word label (.wrd):

7470 11362 she

11362 16000 had

15420 17503 your

17503 23360 dark

23360 28360 suit

28360 30960 in

30960 36971 greasy

36971 42290 wash

43120 47480 water

49021 52184 all

52184 58840 year

Phonetic label (.phn):

(Note: beginning and ending silence regions are marked with h#)

0 7470 h#

7470 9840 sh

9840 11362 iy

11362 12908 hv

12908 14760 ae

14760 15420 dcl

15420 16000 jh

16000 17503 axr

17503 18540 dcl

18540 18950 d

18950 21053 aa

21053 22200 r

22200 22740 kcl

22740 23360 k

23360 25315 s

25315 27643 ux

27643 28360 tcl

28360 29272 q

29272 29932 ih

29932 30960 n

30960 31870 gcl

31870 32550 g

32550 33253 r

33253 34660 iy

34660 35890 z

35890 36971 iy

36971 38391 w

38391 40690 ao

40690 42290 sh

42290 43120 epi

43120 43906 w

43906 45480 ao

45480 46040 dx

46040 47480 axr

47480 49021 q

49021 51348 ao

51348 52184 l

52184 54147 y

54147 56654 ih

56654 58840 axr

58840 61680 h#

6. Online Documentation

-- --------------------

Compact documentation is located in the "/timit/doc" directory. Files in this directory with a ".doc" extension contain freeform descriptive text and files with a ".txt" extension contain tables of formatted text which can be searched programmatically. Lines in the ".txt" files beginning with a semicolon are comments and should be ignored on searches. The following is a brief description of their contents:

phoncode.doc - Table of phone symbols used in phonemic dictionary and

phonetic transcriptions

prompts.txt - Table of sentence prompts and sentence-ID numbers

spkrinfo.txt - Table of speaker attributes

spkrsent.txt - Table of sentence-ID numbers for each speaker

testset.doc - Description of suggested train/test subdivision

timitdic.doc - Description of phonemic lexicion

timitdic.txt - Phonemic dictionary of all orthographic words in prompts

A more extensive description of corpus design, collection, and transcription can be found in the printed documentation.