RT-2002 Evaluation Plan (Version 1.0)

Updated 04/19/2002

1.0 Background

This is the first in a series of Rich Transcription (RT) evaluations being administered by the National Institute of Standards and Technology (NIST). The purpose of this evaluation is to explore the integration of traditional automatic speech recognition (ASR) speech-to-text (STT) generation with automatic metadata (MD) generation over a variety of domains. Traditionally, NIST ASR evaluations have focused on evaluation of single-domain STT technologies in isolation or loosely coupled with the independent generation of a small set of metadata. This evaluation will examine how these technologies can benefit from tighter integration and how they can be made to be flexible to new and different domains.

While there are certainly many possible metadata areas of interest, given the short lead time and limited resources, this evaluation will explore only the automatic segmentation and clustering of speakers within a particular speech excerpt. This will permit NIST and the research community to baseline existing capabilities and begin to build the infrastructure for future enhancements/expansions.

Because of the many possible dimensions for exploration, this evaluation will consist of a variety of evaluated test conditions designed to probe many areas of interest. Since the possible test conditions are many and diverse, no one site will be expected to implement all of them. However, we urge sites to implement all of the test conditions listed as “core”. We also urge sites with complementary capabilities to team-up to create integrated RT systems.

The remainder of this document outlines the data, test conditions, data formats, submission protocols, and schedule for this evaluation.

This document and other documentation, software, and corpora for the RT-02 evaluation are available on the RT-02 website at http://www.nist.gov/speech/tests/rt/rt2002/.

2.0 Evaluation Tasks

This evaluation contains two primary evaluation tasks: Speech-to-Text (STT) generation and Metadata (MD) Annotation.

2.1 Speech-to-Text (STT) Generation Task

The Speech-to-Text generation task for this evaluation consists of generating a stream of lexical tokens from excerpts within audio files. These tokens should be generated in accordance with the rules employed in the 2001 Hub-5 conversational speech recognition evaluation (see below). In short, a word transcription is to be generated for the specified excerpts/channels in the source data using the CTM format required by the SCLITE scoring software.

This year, all recorded events in the test sets must be treated independently. Therefore, sites may not perform any kind of adaptation across events (e.g., news broadcasts, meetings, conversations). However, within-event unsupervised adaptation is permitted and encouraged, including (for certain conditions) making use of multiple recorded channels of an event if they are available.

2.2 Metadata (MD) Annotation Task

The metadata generation task for this evaluation consists of segmenting audio excerpts (files) into speaker changes and clustering these segments by speaker. It will be evaluated in a manner similar to the 2001 Speaker Recognition Evaluation for N-Speaker Segmentation (see below) using the Speaker Segmentation Scoring Software.

This year, all recorded events in the test sets must be treated independently. Therefore, sites may not perform any kind of adaptation across events (e.g., news broadcasts, meetings, conversations). However, within-event unsupervised adaptation is permitted and encouraged, including (for certain conditions) making use of multiple recorded channels of an event if they are available.

3.0. Core Evaluation Conditions

This year, to encourage maximal participation, we are not requiring any particular set of test conditions and sites are permitted to implement either the Speech-to-Text (STT) or Metadata annotation (MD) tasks as they like. However, given that the goal of this evaluation is Rich Transcription, we anticipate that sites will be required to run BOTH STT and MD tasks in the future. We therefore encourage sites with complementary capabilities to team up to create complete RT systems. We also encourage sites to implement STT and MD runs on as many of the data sets/conditions provided as possible to baseline current capabilities – even if they believe that their results will be relatively poor.

3.1 For NSA LVCSR Contractors

NSA LVCSR contractors must include at least one run on the Telephone Conversation test set with speaker segmentation provided (This is the same as last year's Hub-5 test condition). Note that sites should be careful not to look at the speaker segmentation if they are also running the metadata annotation task.

4.0 Supported Evaluation Conditions

This evaluation contains several test dimensions detailed below:

4.1 Processing Speed

Systems may be run at any processing speed, but their category must be specified at the time of submission as either: 1) greater-than-ten-times realtime, 2) less-than-or-equal-to-ten times realtime, or 3) realtime or faster and NIST will sort the results in our tabulations accordingly. We will also request a more specific quantification of your processing speed in your system description.

4.2 Domain

This evaluation contains test corpora from three distinct domains: Broadcast News, Telephone Conversations, and Meetings. Although you are strongly encouraged to implement runs on all three domain types, this year you are permitted to choose to implement recognition a subset of your choosing (unless you have specific contractual obligations).

4.3 Microphone/Channel

The following microphones/channels are supported for each domain:
Broadcast News – single (mono) audio channel
Telephone Conversations – Two distinct audio channels (one for each conversation side)
Meetings -

1) N personal microphone channels (one mic/channel for each meeting participant). The microphone type (head-mount or lapel-mount) will be known.
2) Summed, gain-normalized N personal mics – single mono audio channel
3) Distant omni-directional microphone channel – single mono audio channel

4.4 Segmentation

Manual speaker segmentation is provided for the conversational telephone data for sites who wish to implement the traditional Hub-5 LVCSR task. However, we encourage those sites to also run a segmentation unknown condition as it is likely to be required in future RT evaluations.

As in past Hub-4 evaluations, we will provide the output of the CMU automatic segmenter for the broadcast news data which sites may use if they do not have access to their own segmenter. We do not yet have this capability for the conversational telephone or meeting domain data.

5.0 Training Data

As in previous evaluations, any language corpora/resources which are publicly available at the time of start of the evaluation may be used for training or development testing purposes. Note that there is no specifically-designated development test data for this evaluation. However, the following is suggested training material for this evaluation.

5.1 Broadcast News Training Data

The suggested training data includes all existing Hub-4 and TDT English broadcast news corpora. In the tradition of keeping test/train epochs separate for time-sensitive news data, all news-related material used for training must predate December 1, 1998.

5.2 Conversational Telephone Training Data

The suggested training data consists of the entire Switchboard (i.e., Switchboard I) Corpus as released, the entire Switchboard II phases 1 & 2 Corpus, and all English conversations of the Call_Home Corpus, including those originally designated for training and those used as test data in previous NIST evaluations.

5.3 Meeting Room Training Data

Given the short time frame and that this is a new domain, only a small data set comparable to the evaluation test set will be made available for training. This data consists of 8 recorded meetings with at least a 10-minute excerpt transcribed for each. Three data sets are included for each meeting: individual personal microphone channels, summed personal-mic-mix single audio channel, and distant omni-mic single audio channel. Note that the personal mic data was recorded using either a noise-cancelling head microphone or a lapel microphone.

The speech waveform files for this data are available to test participants on CD-ROM upon request from NIST. The data will be made available to non-participants after the evaluation via the LDC. RT-02-compliant transcriptions of the selected excerpts as well as any transcripts provided by the data collection sites will be made available via anonymous ftp.

6.0 Development Test Data

There is no specific material designated for development testing purposes. As such, the available training material may be partitioned into training and development test sets as desired.

7.0 Evaluation Data

The data for this evaluation is as follows:

7.1 Broadcast News Evaluation Data

The broadcast news test data, approximately 60 minutes in length, will be similar to that used in previous broadcast news evaluations (Hub-4) and consist of single-channel recordings of American news broadcasts. Unlike in previous Hub-4 tests, the test material will be presented in a set of broadcast files and an index file will be provided which specifies specific excerpts for evaluation. No concatenation of excerpts will be performed this year. Note that an entire broadcast may be used for unsupervised adaptation. However, as in the past, cross-broadcast adaptation is not permitted.

7.2 Conversational Telephone Evaluation Data

The telephone data, approximately 300 minutes in length, will be similar to that used in the 2001 conversational speech evaluation (Hub-5). This data will consist of two-channel recordings of whole telephone conversations (one channel for each conversation participant). Each conversation will be provided in a separate file. There will be three test sets drawn from (1) unreleased original Switchboard, (2) Switchboard II phase 3, and (3) Switchboard cellular phase 2. Each test set will contain five minutes from each of twenty conversations, approximately one hour forty minutes in length.

Note: the speaker change/speaker ID information provided for the Hub-5 subtest is not to be used for the speaker metadata annotation task.

7.3 Meeting Room Evaluation Data

This data will consist of recordings of 8 meetings held at CMU, ICSI, LDC, and NIST. An excerpt, approximately 10 minutes long, will be selected for each meeting for evaluation. An index file will be provided to specify the excerpt to be evaluated. The remainder of the recording may be used for unsupervised adaptation. However, cross-meeting adaptation is not permitted. The number of participants in a meeting will range from 3 to 8 and it is possible that some excerpts will contain un-miked participants who “walk in” on the meeting. Also note that some meetings contain non-native speakers. The meetings were collected at several different sites using different data collection equipment/protocols, different subject pools, and different meeting forums.

Three mic/channel types will be used for each meeting and will be presented in separate files:

1) N personal microphone channels (one mic/channel for each meeting participant). The microphone type (head-mount or lapel-mount) will be known.
2) Summed, gain-normalized N personal mics – single mono audio channel
3) Distant omni-directional microphone channel – single mono audio channel

The data from the omnidirectional microphone will be considered “core” and the data from the personal microphones, and personal microphone mix is to be used for contrastive tests. Therefore, a run on the omnidirectional mic data should be included for each test condition implemented.

8.0 Evaluation Software

The software to be used in evaluating the RT-02 results is available from:
http://www.nist.gov/speech/tests/rt/rt2002/evalsoftware.htm

8.1 Speech-to-Text (STT) Evaluation

Sites will generate decodings that include word time alignments and confidence scores. The same scoring algorithm (SCLITE) used for previous Broadcast News evaluations will be used for the scoring of both task for this evaluation. Word error will be the primary metric for all test conditions.

NIST will tabulate and report word error rates over the entire dataset. NIST may also tabulate and report Word Error Rates for various subsets of test material to examine performance for different conditions.

Immediately after the evaluation, NIST will provide the complete annotation record for the evaluation test material, to facilitate the analysis of performance by individual sites.

8.1.1 Orthographic Rules

System developers should familiarize themselves with the orthographic transformations and rules used in preparing both the reference transcriptions and system hypothesis transcriptions prior to official scoring by NIST so that they can obtain the most accurate scoring of their systems.

8.1.1.1 SNOR Format

The transcription format employed for scoring is called SNOR (Standard Normalized Orthographic Representation). The SNOR format provides a common format for recognition output. In doing so, it removes lexical details (such as capitalization, punctuation, etc.) from the transcription format to simplify the recognition and scoring process.

A SNOR-normalized transcription consists of text strings made up of ASCII characters and has the following contraints:

- Whitespace separates words for languages that use words
- The text is case-insensitive (usually in all upper case)
- No punctuation is included except apostrophes for contractions
- Previously hyphenated words are divided into their constituent parts separated by whitespace.

The human-generated reference transcripts are stored in their original detailed format and are translated into SNOR prior to scoring the output of a STT system. It is important that these transformations are properly included in the design of the recognition systems, so that the system output may be scored optimally.

8.1.1.2 Orthographic Normalization

After the reference transcripts are converted into SNOR, both the reference and STT-produced transcripts will be transformed via a NIST supplied transcript filter using a map file. The filter and map file are included with the scoring package on the RT website. Note that the map file is likely to be updated prior to the official scoring, so the final orthographic map file which will be used in the official scoring will not be made available until after the test results are received and scored by NIST. A current version of the filter is made available on the RT website. The sections below on "Multiple Spellings" and "Contractions" describe the mapping process.

8.1.1.3 Notes on the Handling of Special Orthographic Conditions

This section describes orthographic conditions and speech phenomena which require special processing. Note that some of these conditions are scored as "optionally deletable". In these cases, a STT system will not be penalized with an error for omitting the output of (or "deleting") the particular word in the system output. However, reference words marked as such will still count towards the total number of reference words during Word Error Rate (WER) computations.

Word fragments:
Word fragments are represented in the SNOR reference transcription with only the text of what was actually spoken with a "-" indicating the missing portion of the verbally fragmented word. For WER, they are included in the total word count but they are scored as “optionally deletable”. Fragments are also counted as correct if the transcribed portion of the fragment matches the initial substring of an inserted word (e.g. the fragment 'FR-' will be counted as correct if it is aligned to "FRED").

Unintelligible or Semi-Intelligible Words:
Certain words in the reference transcripts may be marked as unintelligible. If the transcriber can make an educated guess at the unintelligible words, a "best guess" may also be provided. These best guesses will be included in the total word count, but will be scored as "optionally deletable".

Foreign Words:
Foreign words, which are outside the language being evaluated, will be annotated as such in the reference transcripts. Such foreign words will be included in the total word count, but will be scored as "optionally deletable". Note that this annotation will not be applied to words of foreign origin that have been widely incorporated into the speech of the evaluated language.

Pause Fillers:
For scoring purposes, all hesitation words, referred to as "non-lexemes", will be mapped to a single form, *%HESITATION* and will be scored as "optionally deletable". Although many different hesitation words are possible (E.g., um, er, uh, ah, etc.), they are all considered to be functionally equivalent from a linguistic perspective. When a hesitation

word is hypothesized, an STT system should emit either nothing or one of the accepted lexical tokens for hesitations. For English, the current list of recognized hesitation words is: *uh, um, eh, mm, hm, ah, huh, ha, er, oof, hee, ach, eee, ew*

Multiple Spellings:
A single standardized spelling will generally be required and the recognizer must output this standard spelling in order to be scored as correct. These spellings are determined by NIST by first consulting the American Heritage Dictionary. If these aren't covered in the AHD, Web searches are performed to find the most common representation. If no single form is most prevalent, then two or more forms will be permitted via the orthographic map file. Spelling errors and multiple representations which commonly occur in the training data will also be permitted via the map file. As in previous years, a transcription filter and orthographic map file will be used on both the reference and hypothesis transcripts to apply rules for mapping common alternate representations to a single scorable form.

Homophones:
Homophones will not be treated as equivalent. Homophones must be spelled correctly according to the given context in order to be counted as correct.

Overlapping Speech:
Periods of overlapping speech will be specially annotated and will not be scored. Any words hypothesized by the recognizer during these periods will not be counted as errors.

Compound Words:
Compound words are divided into their constituent parts in both the ref and hyp unless the word appears as a single word in the American Heritage Dictionary. If the word does not appear in either compounded or un-compounded form in the AHD, then the Web is searched to determine the most likely representation via frequency of occurrence.

Contractions:
Contractions will be transformed into their expanded forms in both the reference and the hypotheses transcriptions. Human annotators will add the proper expansions to the reference transcripts. A list of common contractions with all possible expansions is included in the orthographic map file. The map file will be updated with new contractions occurring in the reference and hypothesis transcripts prior to scoring. The existing approach allows only the "REF" and its expansions for a given contraction to be scored as correct. To implement this, contractions in the recognizer output will be expanded to an alternation of all possible contractions and the alignment routines will select the minimal cost expansion.

8.2 Metadata (MD) Evaluation

The RT-02 evaluation includes only one metadata (MD) evaluation task. Future RT evaluations are likely to support a variety of MD tasks.

8.2.1 Speaker Segmentation and Clustering Evaluation

This task requires two subtasks: 1) segmenting an excerpt of speech into time intervals during which a particular speaker is talking, 2) assigning an arbitrary speaker identifier to intervals from the same speaker. Note that these intervals can overlap, as is the case when two speakers are talking simultaneously. For this task, the unique identity of a speaker is not known and no training material is specified for the speakers in the test excerpt. The number of speakers in a particular excerpt is also unknown.

This task will be evaluated using the speaker segmentation scoring software provided on the RT website. Speaker segmentation will be evaluated by comparing system-determined regions where the various speakers are talking with reference regions determined by NIST. NIST will determine the regions where each speaker is speaking by using time marks determined by human transcribers.

In contrast to previous such evaluations, scoring will be NOT be limited to areas of non-overlapping speech. Note, to avoid mis-scoring tiny time errors in either the reference or hypothesis segmentation, times within 250 milliseconds of the boundary of a reference segment will not be evaluated.

The decisions for each hypothesized speaker will be compared with a reference answer key to determine the segmentation error rates for speaker tracking, according to the following computation:

For the speaker segmentation task, a system must produce hypothesized speaker turns (or segments of speech produced from a particular speaker) and a generic speaker label for each turn (e.g., speaker 0, speaker 1, ...). All turns from the same speaker should have the same generic speaker label. Unlike the other detection tasks, this is a classification task, which can be characterized by a single error rate. (There is only a single error since all speech must be accounted for and a "miss" for one hypothesized speaker label will generate a corresponding "false alarm" for another hypothesized speaker label, so the two errors are no longer independent.) To compute the classification error a search for the best (minimum error) mapping of hypothesized speaker labels to true speakers for each conversation is performed, and then the errors are accumulated over the ensemble of test conversations. The error rate is computed as follows:

Assume a system produces M hypothesized speaker labels for a conversation that actually contains N true speakers. (When the number of speakers is known in advance, we will have $M = N$). When $M < N$ we automatically generate $N\text{-}M$ speaker labels with no associated speech segments. We first produce a one-to-one mapping of true speakers $\{i\}$ with hypothesized speaker labels $\{j\}$

$$map(i) = j; \quad i=1,..,N; \quad 1 <= j <= M$$

Let *true_duration(i, conv)* be the total duration of speech by true speaker i in a conversation denoted *conv*, and let *hyp_duration(i,j,conv)* be the total duration of speech common to true speaker i and hypothesized speaker label j in this conversation. The error rate is then

error(map, conv) = 1
- (S i hyp_duration(i, map(i), conv) /
S i true_duration(i, conv))

The best one-to-one mapping, *map*(i),* used for this conversation is the one producing the minimum *error(map, conv)*. For the conversation we then log the values

hit(conv) = S i hyp_duration(i, map(i), conv)*

and

total(conv) = S i true_duration(i, conv)

The total (weighted) error over an ensemble of conversations is finally computed as
total_error = 1 - S conv hit(conv) / S conv total(conv)

9.0 Speech-to-Text (STT) System Output Format

The STT system output file uses the same CTM format that was used in previous Hub-4 and Hub-5 evaluations. The CTM file format is a concatenation of time mark records for each word in each channel of a waveform. The fields in a record are space-separated;and the records are separated with a newline. Each record must have 1) a waveform id, 2) channel identifier (matching the channels of the reference file), 3) start time, 3) duration, 4) case-insensitive token, and optionally 5) a confidence score. The format is as follows:

<CTM> ::= <F> <C> <BT> <DUR> token [<CONF>]
where

<F> ::= <BNEWS_ID> | <SWB1_ID> | <SWB2_ID> | <SWBCEL_ID> | <MTG_ID>
The waveform id. NOTE: no pathnames or extensions are expected
<BNEWS_ID> ::= bn02en_[1-6]
<SWB1_ID> ::= swDDDD (where DDDD is a four digit conversation code)
<SWB2_ID> ::= sw_DDDDD (where DDDDD is a five digit conversation code)
<SWBCEL_ID> ::= sw_4DDDD (where DDDD is a four digit conversation code)
<MTG_ID> ::= <LOC_ID><MTG_NUM><MIC_TYPE>
<LOC_ID> ::= X (where X is a one letter location code b for recording done at ICSI, c for recording done at CMU, l for recording done at LDC, n for recording done at NIST)
<MTG_NUM> ::= DDD (where DDD is a three digit meeting number)
<MIC_TYPE> ::= t1 (for distance omni directional microphone) | hs (for summed head mics) | hD (for separated head mic where D is the ID of the mic)

<C> ::= the waveform channel. Either "A" or "B" for switchboard, 1 for BNews, or the MIC_TYPE for meeting data.
<BT> ::= the begin time (seconds) of the token, measured from the start time of the wave form.
<DUR> ::= the duration (seconds) of the token
<CONF> ::= optional confidence score

Please refer to *http://www.nist.gov/speech/tests/rt/rt2002/evalsoftware.htm* *for further details.*

9.1 Metadata (MD) System Output Format

The output generated by a Speaker Segmentation system is formatted as a list of segmentation records separated by SGML tags indicating the audio file used to derive the segmentation. Each segmentation record indicates the start and end times for a speaker, and the speaker id to which the interval is attributed. The file is formatted as follows:

<**segment filename**=SEGMENT_NAME>
START_TIME END_TIME SPEAKER_ID
START_TIME END_TIME SPEAKER_ID
</**segment filename**>

where

START_TIME ::= the starting interval time (to the hundredth of a second).
END_TIME ::= the ending interval time (to the hundredth of a second).
SPEAKER_ID ::= the speaker cluster this segment belongs to. This can be any string.

Please refer to *http://www.nist.gov/speech/tests/rt/rt2002/evalsoftware.htm* *for further on how to evaluate a system.*

10.0 Submission Instructions

All results must be submitted by Friday April 19, 2002. Sites should submit results using the following steps:

Step 1: Directory Structure

Create a directory identifying your site ('SITE') which will serve as the root directory for all your submissions. Examples:

- bbn
- dragon
- ibm
- sri
- …

You should place all of your test results in this directory. When scored results are sent back to you and subsequently published, this directory name will be used to identify your organization.

For each test system, create a sub-directory ('SYS') under your 'SITE' directory identifying the system's name or key attribute and evaluation conditions. The sub-directory name is to consist of a string containing 1) a system ID chosen by you, 2) the task for which the system was used (Speech-to-Text or MetaData), 3) the domain of the task (Broadcast News, Telephone Conversations, or Meetings), 4) the speed of the system (<=1xRT, <=10xRT, >10xRT), 5) the microphone/channel that the system used, 6) the type of segmentation the system used. Place all files pertaining to the tests run using a particular system in the same sub-directory. The following is the BNF directory structure format:

<SITE>/<SYS>/<FILES>
where
    <SITE> ::= bbn | dragon | ibm | sri | . . .
    <SYS> ::= <SYS_ID>_<TASK>_<DOMAIN>_<SPEED>_<MIC>_<SGN>
    where
        <SYSID> ::= (short system description ID, preferably <= 8 characters)
        <TASK> ::= stt | md
        <DOMAIN> ::= bnews | swbd | mtg
        <SPEED> ::= le1xRT | le10xRT | gt10xRT
        <MIC> ::= 1 | 2 | n | sum | omni
            "1" for Broadcast news, "2" for switchboard, "n" for the N meeting data personal mics, "sum" for the meeting data summed personal mics, and "omni" for the meeting room central omin mic.
        <SGN> ::= manual | auto
    <FILES> ::= sys-desc.txt | <TEST_SET>.ctm | <TEST_SET>.sgn
    where
        sys-desc.txt ::= system description, including reference to paper if applicable (see Step 3 for further details)
        <TEST_SET>.ctm ::= file containing time-marked hypothesis token.
        <TEST_SET>.sgn ::= file containing time-marked hypothesis segment.
        where
            <TEST_SET> ::= base name of the corresponding UEM or PEM file.

Step 2: System Output Files
The system outputs must follow the format given in Sections 9.0 and 9.1.

To prepare a STT result, concatenate all CTM records for a test set, as identified by the UEM/PEM file name, into a single file, <TEST_SET>.ctm, and place it in the <SYS> directory,

To prepare a MetaData submission, construct your system output as specified in Section 9.0 and place the results for a single test set, as identified by the UEM/PEM file name, into a single file, <TEST_SET>.sgn and place it in the <SYS> directory.

Step 3: System Documentation
For each test you run and for each system evaluated, a brief description of the system (the algorithms) used to produce the results must be provided along with the results. (It is permissible for a single site to submit multiple systems for evaluation. In this case, however, the submitting site must identify one system as the "primary" system prior to performing the evaluation.) The format for the system description is as follows:
SITE/SYSTEM NAME
TEST DESIGNATION

1. Primary Test System Description:
2. Acoustic Training:

3. Grammar Training:
4. Recognition Lexicon Description:
5. Differences for each Contrastive Test: (if any contrastive test were run.)
6. New Conditions for This Evaluation:
7. Execution Time:
    - Sites must report the CPU execution time that was required to process the test data, as if the test were run on a single CPU. Sites must also describe the CPU and the amount of memory used:
8. References:

Step 4: Submission Protocol
Once you have structured all of your recognition results according to the above format, you can then submit them to NIST via ftp. The following instructions assume that you are using the UNIX operating system. If you do not have access to UNIX utilities or ftp, please contact NIST to make alternate arrangements.

Ftp method:
First change directory to the directory immediately above the <SITE> directory. Next, type the following command.

tar -cvf - ./<SITE> | compress > <SITE>-<SUBM_ID>.tar.Z

where

<SITE> is the name of the directory created in Step 1 to identify your site.
<SUBM_ID> is the submission number (e.g. your first submission would be numbered '1', your second, '2', etc.)

This command creates a single file containing all of your results. Next, ftp to jaguar.ncsl.nist.gov giving the username 'anonymous' and your e-mail address as the password. After you are logged in, issue the following set of commands, (the prompt will be 'ftp'):

- ftp> cd /incoming/
- ftp> binary
- ftp> put <SITE>-<SUBM_ID>.tar.Z
- ftp> quit

You've now submitted your recognition results to NIST. The last thing you need to do is send an e-mail message to Audrey Le at 'audrey.le@nist.gov' and copy Jonathan Fiscus at 'jonathan.fiscus@nist.gov' notifying NIST of your submission. Please include the name of your submission file in the message.

Note:

If you choose to submit your results in multiple shipments, please submit ONLY one set of results for a given test system/condition unless you've made other arrangements with NIST. Otherwise, NIST will programmatically ignore duplicate files.

11.0 Schedule

2002	***Event***
Jan. 02	Initial Announcement
Jan. 14	NIST posting sample data for "rich transcription" experiment
Jan. 31	Annotated sample data from sites due at NIST
Mar. 15	Initial meeting room training data released
Mar. 15	Evaluation software released
Mar. 18	Evaluation Plan released
Mar. 22	Additional meeting room training data released
Mar. 25	Commitment deadline for participation
Apr. 01	Evaluation data released
Apr. 19	Results due at NIST 2400 EDT
Apr. 26	Official results released by NIST
Workshops	
May 7-8	RT-02 Evaluation Workshop
May 9-10	EARS Kickoff Meeting