

NIST Rich Transcription Evaluation

Spring 2005 Evaluation

Registration Agreement Form

To participate in the Spring 2005 NIST Rich Transcription Evaluation (RT-05S) this registration form must be filled out in its entirety and returned to NIST by the registration deadline of 28-April-2005.

FAX: Attention Jonathan Fiscus
+1 (301) 670-0939

or

Postal Mail: Jonathan Fiscus - NIST
100 Bureau Dr. Mail Stop 8940
Gaithersburg, MD 20899-0940 USA

The following are terms of participation in the RT-05S evaluation:

- On receiving the evaluation data (12-May-2005), the registered site will process the entire test set in accordance with the RT-05S Evaluation Plan, and return the system output files to NIST by the end of the evaluation period (26-May-2005 5:00 pm EDT).
- The site agrees to be represented at the MLMI 2005 Meeting Recognition Workshop to be held in Edinburgh, UK 13-July-2005, and to openly discuss the system(s) evaluated.
- The evaluation data cannot be redistributed. Sites may keep the data and continue using the data for research purposes.

The following are terms of publishing the NIST Rich Transcription evaluation results:

- The site understands that, although the evaluation is open to all wishing to participate, results are only discussed within the group of registered participants and representatives of supporting government agencies.
- The site agrees to not publicly compare its results with the results of other participants, such as stating rank ordering. Sites are free to do what they wish with their own results, but may neither redistribute nor publish results from another site without that site's explicit permission.

Failure to abide by these terms jeopardizes a site's chance to participate in future NIST evaluations.

NIST serves to coordinate the NIST Rich Transcription evaluations. These tests are designed for local implementation by each participant. Only their results are returned to NIST for scoring. The reported results are not to be construed, or represented, as endorsement of any participant's system, or as official findings on the part of NIST or the U.S. Government.

Site Affiliation: _____

Evaluation Tasks: (check boxes of planned participation)

	Conference Room Data	Lecture Room Data
Speech-To-Text		
Speaker Diarization		
Speech Activity Detection		
Source Localization		

Contact Information

Name: _____ Address for evaluation data: _____

Phone: _____

Signature: _____

Date: _____

By signing and returning this form to NIST you are registering your site as a participant in the NIST RT-05S evaluation, and acknowledging that you have read and will abide by the protocols as described here and in the official evaluation plan.

NIST Rich Transcription Evaluation

Spring 2005 Evaluation

Data Availability Addendum

This form identifies the corpora available to system developers for the Spring 2005 NIST Rich Transcription Evaluation (RT-05S). After signing and returning this form along with the RT-05S evaluation agreement, the data distribution agencies will be notified of the participant's eligibility to receive the data. While some of the resources below are publicly available through the Linguistic Data Consortium (LDC) and/or the Evaluations and Language resources Distribution Agency (ELDA), others below are only distributed through special arrangements such as this evaluation. Participants agree to use all corpora in accord with the licenses which accompany each resource.

Publicly available meeting resources:

- ICSI Meeting Speech: LDC catalog number LDC2004S02
- ICSI Meeting Speech: LDC catalog number LDC2004T04
- ISL Meeting Speech Part1: LDC catalog number LDC2004S05
- ISL Meeting Transcripts Part 1: LDC catalog number LDC2004T10
- NIST Meeting Pilot Corpus Speech: LDC catalog number LDC2004S09
- NIST Meeting Pilot Corpus Transcripts and Metadata: LDC catalog number LD2004T13
- Rich Transcription 2004 Spring (RT-04S) Development & Evaluation Data

Non-publicly available corpora offered to the RT-05S evaluation participants:

The corpora listed in this section have been produced by several non-affiliated programs. A data sharing agreement has been reached whereby sites not affiliated with each corpus' producer are granted a non-transferable evaluation license to the data. Sites are allowed to retain and use the data for research purposes.

- Computers in the Human Interaction Loop (CHIL) development test set: a five meeting data set collected by the CHIL Consortium and distributed to non-CHIL partners as a resource for developing RT-05S systems.
- Augmented Multiparty Interaction (AMI) development test set: a twelve meeting data set collected by the AMI project and distributed to non-AMI partners as a resource for developing RT-05S systems.
- Effective, Affordable, Reusable, Speech-To-Text (EARS) RT-04 Broadcast News training corpus distributed to non-EARS partners as a resource for developing RT-05S systems.
 - Topic Detection and Tracking Phase 4 (TDT4) corpus: LDC catalog number LDC2003E02
- Effective, Affordable, Reusable, Speech-To-Text (EARS) RT-04 Conversational Telephone Speech training corpus distributed to non-EARS partners as a resource for developing RT-05S systems:
 - Fisher English Training Speech Part 1 Speech: LDC catalog number LDC2004S13 (5850 two sided telephone conversations)
 - Fisher English Training Speech Part 1, Transcripts: LDC catalog number LDC2004T19 (5850 transcribed two sided telephone conversations)
 - Fisher English Training Speech Part 2: LDC catalog number LDC2005S13
 - Fisher English Training Speech Part 2, Transcripts: LDC catalog number LDC2005T19

By signing this form, the site acknowledges they have returned to NIST a completed RT-05S Evaluation Agreement form and agrees to the terms of use for the aforementioned non-publicly available resources.

Site Affiliation: _____

Signature: _____

Date: _____

LDC Evaluation Resources Agreement for RT-05S

In the remainder of this document the term **User** refers to:

_____ (Individual name)

and the term **User's research group** refers to:

_____ (University, Institute or Company name)

_____ (Specific department or area, if appropriate).

This document describes the terms of an agreement between **User** and the Linguistic Data Consortium (LDC), in which **User** will receive material as specified below.

Under this agreement, **User** will receive media (CD-ROM, DVD, hard drive, etc.) containing speech and/or text data named below under "Corpora Received." **User** agrees to use the material received under this agreement only for non-commercial linguistic education and research purposes. In the event that **User's** use of LDC Corpora results in the development of a commercial product, **User** must join the LDC as a Commercial Member and pay all applicable fees prior to release of said commercial product. Unless explicitly permitted herein, **User** shall have no right to copy, redistribute, transmit, publish or otherwise use the LDC Corpora for any other purpose and **User** further agrees not to disclose, copy, or re-distribute the material to others outside of **User's research group**.

User agrees that neither the Linguistic Data Consortium nor the University of Pennsylvania warrant the accuracy, completeness, currentness, merchantability or fitness for a particular purpose of the information. In no event will the Linguistic Data Consortium or the University of Pennsylvania be liable to any authorized user, or anyone else for any loss or injury caused in whole or in part by its negligence or contingencies beyond its control in procuring, compiling, interpreting, editing, writing, reporting or delivering the information, or any errors, omissions or inaccuracies in the information, regardless of how caused. In no event will the Linguistic Data Consortium or the University of Pennsylvania be liable to the organization, any authorized user or anyone else for any decision made or action taken by the organization or any authorized user in reliance upon any part of the information or for any consequential, direct, special or similar damages, even if advised of the possibility of such damages.

Corpora Received:

- ICSI Meeting Speech: LDC catalog number LDC2004S02
- ICSI Meeting Speech: LDC catalog number LDC2004T04
- ISL Meeting Speech Part1: LDC catalog number LDC2004S05
- ISL Meeting Transcripts Part 1: LDC catalog number LDC2004T10
- NIST Meeting Pilot Corpus Speech: LDC catalog number LDC2004S09
- NIST Meeting Pilot Corpus Transcripts and Metadata: LDC catalog number LD2004T13
- Topic Detection and Tracking Phase 4 (TDT4) corpus: LDC catalog number LDC2003E02

LDC Evaluation Resources Agreement for RT-05S (page 2)

- Fisher English Training Speech Part 1 Speech: LDC catalog number LDC2004S13
- Fisher English Training Speech Part 1, Transcripts: LDC catalog number LDC2004T19
- Fisher English Training Speech Part 2: LDC catalog number LDC2005S13
- Fisher English Training Speech Part 2, Transcripts: LDC catalog number LDC2005T19

Organization: _____

Name: _____

Signature: _____

Date: _____

For LDC:

Christopher Cieri
Executive Director

AMI Evaluation Resources Agreement for RT-05S

The Augmented Multiparty Interaction (AMI) Program will supply participants in the Spring 2005 Rich Transcription Meeting Recognition (RT-05S) Evaluation with meeting recordings collected at the Institute for Communicating and Collaborative Systems, School of Informatics, The University of Edinburgh and the Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP). All recordings and ancillary information described in the Corpora section below and any other data distributed by the AMI program to the RT-05S participants are licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.0 agreement. Under this agreement, licensees have the following rights.

Summary of the Creative Commons

Attribution-NonCommercial-ShareAlike 2.0 Agreement

(The full Legal Code is at <http://creativecommons.org/licenses/by-nc-sa/2.0/legalcode>.)

You are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- **Attribution.** You must give the original author credit.
- **Noncommercial.** You may not use this work for commercial purposes.
- **Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.
 - For any reuse or distribution, you must make clear to others the license terms of this work.
 - Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

Corpora

The AMI program will make available the following corpora to the RT-05S community.

- Two meetings for the evaluation corpus
- Three sets of four meetings to be used as development test set
- Additional meetings may be supplied by the AMI program and will be covered by this agreement.

By signing this document, the undersigned agrees to use the AMI corpora resources in accord with the Creative Commons Attribution-NonCommercial-ShareAlike 2.0 agreement.

Organization: _____

Name: _____

Signature: _____

Date: _____