

# NIST-DARPA Interagency Agreement: SLS Program

*David S. Pallett, Principal Investigator*

National Institute of Standards and Technology  
Building 225, Room A-216  
Gaithersburg, MD 20899

## Project Goals

1. To provide a central role in the design, development and distribution of speech and natural language corpora for the DARPA SLS research community.
2. To design, coordinate implementation of, and analyze results of performance assessment tests for DARPA's Spoken Language Systems.
7. In collaboration with Texas Instruments (and using NIST funding), prepared the TI Connected Digits Database for production on a three-CD-ROM set (CD4-1.1, 4-2.1, 4-3.1).

## Plans for the Coming Year

1. Complete production of the ATIS0 six-CD-ROM set (adding CD5-5.1, and CD5-6.1), incorporating corrections to the documentation and transcription files as possible.
2. Produce ATIS material provided by SRI (and others, as resources permit) on CD-ROM in "standard" format with auxiliary files. Additional material may include corpora collected to date by CMU and MIT.
3. Collaborate with others (e.g., CMU, and possibly TI and/or NYNEX) in preparing other speech corpora for release in CD-ROM format.
4. Participate in Performance Evaluation Working Group activities within the DARPA SLS Program.
5. Analyze the results of the February 1991 Benchmark Tests.
6. Prepare for the October 1991 Benchmark Tests.

## Recent Results

1. Prepared the complete DARPA TIMIT Corpus for release on CD-ROM (CD1-1.1). This release includes the complete 630-speaker corpus, verified/corrected time-aligned phonetic labels, new time-aligned word-boundary labels, and a phonemic lexicon.
2. Prepared the ATIS0 Pilot Corpus for release in CD-ROM format. Two single discs (CD5-1.1, 5-2.1), and two (CD5-3.1, 5-4.1) (out of a planned set of six) discs have been prepared and released within the DARPA Community.
3. Prepared the DARPA Spoken Language Systems February 1991 Benchmark Test Material Disc (CDT-1.1).
4. Defined and developed additional test procedures and software for evaluating spontaneous ATIS SPeech RECOgnition (SPREC) systems, and for ATIS Class D1 query pairs.
5. Coordinated DARPA Corpora and Performance Evaluation Committee and Working Group Activities.
6. Administered and scored the February 1991 DARPA SLS Resource Management and ATIS Benchmark Tests.