

Workshops on Large Vocabulary Conversational Speech Recognition at Johns Hopkins

Frederick Jelinek

Johns Hopkins University
Baltimore, MD, 21218
jelinek@jhu.edu

ABSTRACT

This article describes the organization and plans of the DoD Workshops on Conversational Speech Recognition at Johns Hopkins

1. The LVCSR workshop tradition

In 1991 Jordan Cohen decided to organize workshops concerned with the problem of transcription of conversations conducted over the telephone. He received the backing of Curt Boylls, and together they assembled the needed support. They persuaded Jim Flanagan to become the host at CAIP/Rutgers of the first two summer workshops: 1993 and 1994. The three together organized and carried out both.

When I became the director of the Johns Hopkins Center for Speech Processing¹ it was suggested that future workshops be held at our Center. I agreed and two more workshops took place in the summers of 1995 and 1996. We are now beginning to plan the 1997 workshop.

2. The planning process for LM95

In July and August 1994 we met at Rutgers with many interested parties, most (but not all) of whom were workshop participants at the time, and decided that the subject of the 1995 workshop would be language modeling, that the work would be carried out by four teams, each attacking a separate aspect, that one of the projects would concern Spanish, and that success would be measured by error rate attained on standard test corpora [1] and [2].

We set out to organize along these lines. The next meeting took place in November 1994 in a motel in Bowie, MD at the occasion of one of the LVCSR NSA meetings. We settled on projects and project leaders:

- **Spanish** (Herb Gish, BBN)
- **Language modeling for spontaneous speech** (Roni Rosenfeld, CMU)
- **Fast training and portability** (Mitch Weintraub, SRI)
- **Phrase structure language models** (Salim Roukos, IBM)

We assembled a list of potential participants and agreed to aim at 5–

¹This was its name at the time. It is now the Center for Language and Speech Processing.

member teams, each to include a student and a DoD researcher. We hoped to invite several non-North Americans as participants. Each team was to be helped by one Johns Hopkins student made familiar with the computing environment. The teams were essentially set by the end of 1994. We noted with pleasure that practically all of our invitations were accepted.

Here are some organizational and planning details of interest:

- Each participant was provided with a workstation, either a SUN Spark 5 or an SGI Indigo 2 with 256 MB of memory. Available was also a dual processor SGI Challenge Series L server.
- Housing was in apartments in a building within walking distance from the campus.
- Two special weeks were organized with dedicated topics and invited experts:
 - Linguistics week: Groenendijk, Kroch, Mac Donald, Seidenberg.
 - Information Theory week: Cover, Csizsar, Narayan, Ziv.
- Eight outside lecturers spoke on a variety of topics of interest.
- Every week there was held a two hour meeting at which the four teams reported on their progress.
- The last two days of the workshop were dedicated to reports on results.

For those interested, a rather complete set of reports on many aspects of LM95 can be found in [3], [4], [5], and [6]. All the programs and data developed by and for LM95 are available through the CLSP web page <http://www.clsp.jhu.edu>.

3. The planning process for WS96

In order to select research topics for the 1996 workshop in a really thorough manner, we organized a special two and a half day meeting in November 1995 at the Conference Center in Airlie, PA. We invited to it 25 international participants. The following projects were decided upon (project leaders in parentheses):

- **Speech data modeling** (Jordan Cohen, IDA)
- **Automatic learning of word pronunciation from data** (Mitch Weintraub, SRI)

- **Modeling systematic variations in pronunciation via a language-dependent speaking mode** (Mari Ostendorf, BU)
- **Non-local dependency modeling using link grammars** (Andreas Stolcke, SRI)

Certain new workshop practices were agreed on:

- Each team would meet face to face to settle details about the approach to be taken and to enable intensive workshop preparation during the Spring of 1996. In this way actual research would really *begin* on July 15.²
- Every team would have two students, one of whom would be present for 8 weeks to assure smooth software installation and running.³
- There would be funds to support follow-up work by students; selection of recipients by independent committee.
- Team leaders themselves would be responsible for assuring that appropriate software facilities and data are provided.
- Special weeks (e.g., Information Theory or Linguistics) would be dispensed with and there would be at most one outside speaker per week.

As intended, preparation for WS96 was much more intensive than before. Fortunately, Lin Chase (CMU) agreed to be the year-round coordinator of these efforts. As before, most invitations for participation were accepted. Reports on results of WS96 will again be available as will all the programs and data developed by and for WS96 (through the CLSP web page <http://www.clsp.jhu.edu>).

4. Benefits of LVCSR Workshops

The recognition results of the past two workshops were amply described in the preceding presentations. Here is a sampling of what are, in my opinion, the most important technical benefits:

LM95

- SRI follow-up on difference in recognition performance on simulated and real conversation
- The idea of bleaching as a means to integrate language modeling information from diverse domains
- Thorough recognition performance analysis and benchmarks
- SWITCHBOARD established internationally as a recognition task of interest

WS96

- Non-local dependency language modeling
 - Maximum entropy toolkit
- Automatic learning of word pronunciation from data
- Speaking mode – dependent dynamic pronunciation
- Multi-channel acoustic processing
 - Acceptance of hybrid HMM-ANN technology by the mainline community
- General eagerness to carry out follow-up work

²Actually, several teams met twice prior to WS96

³The demand for student slots was enormous. In some instances we yielded to the demand and one team even ended up with 4 students.

- Research of 3 students will be supported
- Two STIMULATE proposals were submitted[7]

Of course, the workshop brings about many less tangible benefits that may prove much more important in the long run. First among these is the education of the student interns. Next is the cross-fertilization of ideas between disparate industrial, university, and government participants that is the result of the open atmosphere and the attendant esprit-de-corps. The great variety of initial exploratory experiments provides a rich source of information for future research. Software tools are developed and data is refined and collected, both ready for later use by the entire field. Finally, the entire process starting with the discussion leading to problem selection, continuing with the great effort at the workshop itself, and ending with a vigorous follow-up provides an invaluable push for our technology.

5. Future workshops

Our intentions are for the workshops to become a place where an ever widening international circle of researchers cooperates to improve the LVCSR state-of-the-art, where initiated students become integrated into the community, and where real innovation *begins*.

It is in the spirit of this goal that we are starting to plan for WS97. Besides further intensifying the preparatory process, we intend to reach out to so far uninvolved international colleagues, and to provide increased opportunity for student participation.

6. REFERENCES

1. J.J. Godfrey, E.C. Holliman, and J. McDaniel: "SWITCHBOARD: Telephone speech corpus for research and development," *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp 517 - 520, Detroit, 1995
2. C. McLemore: "1995 - 96 LDC Callhome Effort," *Proceedings of the LVCSR Workshop*, Maritime Institute of Technology, Linthicum, MD, October 30 - 31, 1995
3. *1995 Language Modeling Summer Research Workshop: Opening and Closing Day Reports*, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, 1995.
4. *1995 Language Modeling Summer Research Workshop: Proceedings*, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, 1995.
5. *1995 Language Modeling Summer Research Workshop Technical Reports*, Research Notes No.1, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, February 27, 1996.
6. F. Jelinek: *Department of Defense Workshops on Conversational Speech Recognition at Johns Hopkins University*, Research Notes No.6, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, February 27, 1996.
7. *Stimulate*, an NSF program of grants in Information Science, September 1, 1996.