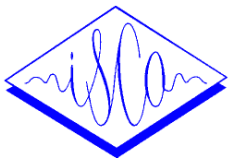

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

NEW TELEPHONE SPEECH CORPORA At CSLU

R. A. Cole *M. Noel* *T. Lander* *T. Durham*

email: cole@cse.ogi.edu http://www.cse.ogi.edu/CSLU/
Center for Spoken Language Understanding, Oregon Graduate Institute
P. O. Box 91000, Portland, Oregon 97291-1000 USA

ABSTRACT

The Center for Spoken Language Understanding (CSLU) collects, annotates and distributes telephone speech data to enable research in spoken language understanding and automatic language identification. This paper gives a brief overview of recent activities in pursuit of this mission. We summarize corpus development activities at CSLU and describe new corpora useful for research on specific tasks: alphabet recognition, numbers recognition, large vocabulary word recognition, and yes/no recognition. We then discuss our two newest data collection efforts, Cellular Speech and the 22-Language Telephone Speech Corpus. All CSLU corpora are available at no charge to academic institutions.

1. OVERVIEW OF CSLU CORPUS DEVELOPMENT ACTIVITIES

Corpus development activities at CSLU include:

(a) **Protocol development**
(b) **Data collection**
(c) **Development of tools**
(d) **Transcription**
(e) **Convention development and documentation**
(f) **Reliability studies**
(g) **Distribution of data**

Protocol Development One of the first steps in any data collection is the development of a protocol that will elicit responses appropriate for the kind of systems one is planning to build. We design our protocols in a variety of ways, but we maintain a focus on continuous or "natural" telephone speech. In addition to our continuous speech corpora, we have various large corpora containing repeated or isolated words and phrases, spoken letters names and numbers. See Table 1 for a list of all speech corpora at CSLU.

Data Collection Once the protocol is determined, it is necessary to create an automatic system to answer the telephone and record each caller's responses. Our system is accessible via a toll-free number throughout the United States. This not only increases our subject base, but also decreases possible dialectical bias. In addition to English data collections, we are collecting speech in 21 other languages as a part of our 22 Language Corpus.

Transcribing Speech We generally have a staff of 5-10 trained transcribers who label at various levels. We produce transcriptions aligned to the waveform, both word and phonetic levels, as well as word transcriptions not aligned to the waveform. All transcriptions explicitly capture any extraneous noise in the signal, such as breath noise or background speech. Pauses, or periods of relative silence in the signal, are also marked.

Convention Development CSLU develops and documents all transcription conventions used in the transcribed speech corpora. For complete coverage of current CSLU labeling conventions, contact Terri Lander at tlander@cse.ogi.edu, or see [1] and [2].

Labeler Reliability We periodically run experiments measuring inter-labeler reliability. This provides a baseline for inter-labeler agreement which can then help us realistically mea-

sure system performance. These studies also validate the quality of data transcribed at CSLU. [3]

Distribution Finally, in keeping with the mission of CSLU, to advance the state of the art in spoken language systems, we make all of our corpora available at no charge to academic institutions as well as to our members. For information on obtaining any corpora, contact Mike Noel at noel@cse.ogi.edu. The CSLU speech tools [4] are also in the public domain.

2. CURRENT STATUS OF CSLU CORPUS DEVELOPMENT

Table 1 lists all of the corpora that the Center for Spoken Language Understanding has completed or is currently developing. The descriptions contained in the table are very brief. For more extensive information, contact Mike Noel at noel@cse.ogi.edu, see publications [5] and [6], or search the online documentation on the Internet: http://www.cse.ogi.edu/CSLU/.

The "Status" column in Table 1 will contain the number of the most current release. If the corpus has not yet been released the status column will contain the expected release date.

Table 1: *Speech Corpora at CSLU including a brief description, availability and current status.*

Corpus	Description	Status
ISOLET	Isolated letters	1.0
Spelled and Spoken Names	Spelled and spoken names	1.0
OGI-TS	11 languages	1.0
Stories	Continuous English speech	2.0
Cellular Words and Phrases	Cellular speech	1.0
Apple Words and Phrases	Repeated phrases	1.0
Alphabets and Spelled Words	Fluent letters	12/95
30K Numbers	Fluent numbers	1.0
30K Names	Fluent names	1.0
Yes/No Corpus	Yes/no utterances	12/95
22 Language Corpus	22 languages	in prog
Cellular Speech	Cellular speech	in prog

3. NEW CORPORA

During the past five years, CSLU has collected speech data for many different projects. We have now undertaken a major effort to consolidate speech from these different data collection efforts into separate corpora that will be of use to researchers and developers. These corpora include alphabets and spelled words, spoken names, numbers, and yes/no responses.

3.1. Alphabets and Spelled Words

This new corpus consists of spelled names and recitations of the English alphabet. The Alphabets and Spelled Words Corpus is taken from several different corpora, including Spelled and Spoken Words [7], Census [7], Cellular Words and Phrases [5] and Cellular Speech, described below. Recitations of the English alphabet are taken from the Spelled and Spoken Words Corpus and Cellular Words and Phrases. All files are transcribed at the word level, and a portion of the files are transcribed phonetically. Our goal for this corpus is to collect 20,000 spelled names and alphabets, and transcribe 2,000 phonetically.

Currently the Alphabets and Spelled Words Corpus consists of 5,114 files, 742 of which are phonetically transcribed. The first release is expected to be available by the end of 1995.

3.2. 30,000 Numbers

The 30K Numbers Corpus consists of over 30,000 numbers utterances. These utterances are taken from our previous data collections, just as for the Alphabets and Spelled Words Corpus. Any phrase that can be considered a number is placed in the corpus: cardinal numbers, ordinal numbers, and digit strings.

Each file in this corpus has an orthographic transcription and a large number of the files have a phonetic transcription. Phonetic transcription is done in order to maximize the coverage of different phonetic contexts.

Version 1.0 of this corpus has been released.

This release consists of 15,000 utterances. Of these, 6,595 have been phonetically transcribed.

3.3. 30,000 Names

The 30,000 Names Corpus is a collection of examples of speakers saying names. Name utterances, which can be first names, last names, middle names, or initials, are taken from utterances from previous data collections. Each file has an orthographic transcription.

Phonetic transcription is done in order to maximize the coverage of different phonetic contexts. Because of the phonetic variability of names in English, a wide variety of phonetic contexts are represented, making this ideal for training vocabulary independent continuous speech recognizers.

Version 1.0 of this corpus has been released. This release consists of 15,000 utterances. Of these, 6,318 have been phonetically transcribed. With these transcriptions tokens of approximately 40 of all possible phonemic bigrams have been transcribed.

3.4. Yes/No

The Yes/No Corpus consists of approximately 80,000 examples of speakers saying "yes" or "no". We also include a small variety of other affirmative and negative responses, such as "yep" and "nope".

The yes and no utterances are responses from questions similar to the following:

- "Have you ever been married? Please say yes or no."
- "Are you of Spanish or Hispanic origin?"
- "Are you calling from within a vehicle?"

Each file in the corpus has an orthographic transcription and a large number of the files have phonetic transcriptions.

Currently the corpus consists of 12,177 files. No manual phonetic transcription has been started yet, although utterances have been aligned automatically and spot checked for accuracy.

The first release is expected to be available by the end of 1995.

4. NEW DATA COLLECTIONS

In addition to the consolidated corpora described above, in which data are combined from previous efforts, CSLU is now engaged in two new data collection efforts, Cellular Speech, and the 22 Language Corpus. The latter corpus is a massive new effort to collect speech in many different languages. A paper describing it is is being submitted separately to EUROSPEECH95.

The second new data collection is being developed in conjunction with a major cellular telephone service provider. This data collection is the Cellular Telephone Speech data collection. The resulting corpus will contain files from 5000 different talkers, each calling from a cellular telephone. Each speaker will provide examples of several different kinds of utterances, including fixed vocabulary utterances (e.g., *male or female*) and spontaneous speech.

All of the files will be transcribed at the non-time-aligned word-level.

4.1. Cellular Protocol

The protocol is divided into two different sections. In the first section the participant provides information about the environment and other factors that may affect the acoustics of the signal. The second section collects examples of names, numbers, fixed vocabulary utterances, and spontaneous speech.

We can not assume that callers are speaking from within a car, so we have two different protocols with slightly different questions. For instance, we don't ask *"About how fast are you traveling right now?"* for callers not in a vehicle. Below we list the questions asked callers within a vehicle:

- The first four questions provide us with background information.
  1. Are you male or female?
  2. What is your native language?
  3. What city and state did you grow up in?
  4. What is your date of birth?
- The answers to the next set of questions will provide information about driving conditions and the recording conditions in caller's car.

1. Please tell us if your window is open, or if you are using the windshield wipers, heater or radio.
2. Briefly describe the traffic conditions.
3. About how fast are you traveling right now?
4. Are you using a digital or analog phone?
5. If you know the brand and model of your cellular phone, please tell us now.
6. Are you using your phone's handset or a mounted microphone?

- The answers to the next questions provide us with some background information about the caller and some examples of spoken digits and letters.

1. Please say your last name.
2. Please spell your last name.
3. Please say a familiar license plate number.
4. Please say a familiar phone number.
5. What time is it now?
6. Please say another phone number.
7. What is today's date?
8. Please say the days of the week.

- The last question is designed to provide samples of natural continuous speech. It contains up to 30 seconds of continuous speech. We suggest, but do not limit the caller to the following topics:

1. something about self
2. a typical day
3. positive aspects of living area
4. family
5. dream home
6. home town
7. favorite restaurant
8. favorite sport or hobby
9. favorite movie or television show

When the phone call is completed, each caller is offered a gift certificate to various national food or retail chains in appreciation for the call.

4.2. Transcription

Each file in this corpus is listened to by a human verifier who makes judgements as to the completeness and appropriateness of the answer. Orthographic transcriptions of each utterance are created.

4.3. Current Status

Currently we have 365 complete calls collected in the Cellular Telephone Speech Corpus. The first release of this corpus will be available by the end of 1995.

5. REFERENCES

[1] T. Lander, S.T. Metzler, The CSLU Labeling Guide, CSLU, Oregon, February, 1994.
[2] J.L. Hieronymus. ASCII phonetic symbols for the world's languages: Worldbet. AT&T Bell Laboratories Technical Memo, 1994.
[3] T. Lander, B.T. Oshika, R.A. Cole, and M. Fanty. Multi-language Speech Database: Creation and Phonetic Labeling Agreement. *Proceedings of the International Congress of Phonetic Science.* Stockholm Sweden, August 1995.
[4] CSLU. "The OGI Speech Tools User's Manual,: Technical report, Center for Spoken Language Understanding, Oregon Graduate Institute, 1993.
[5] Cole, R. A., M. Fanty, M. Noel and T. Lander, "Telephone Speech Corpus Development at CSLU", Proceedings of the 1994 International Conference on Spoken Language Processing, Yokohama, Japan, Sept. 18-22, 1994.
[6] Cole, R. A., M. Noel, D. C. Burnett, M. Fanty, T. Lander, B. Oshika and S. Sutton, "Corpus development activities at the Center for Spoken Language Understanding," Proceedings of the ARPA Workshop on Human Language Technology, April 7-11, 1994.
[7] R. A. Cole, K. Roginski, and M. Fanty, "A Telephone Speech Database of Spelled and Spoken Names", Proceedings of the International Conference on Spoken Language Processing, Banff, Alberta, Canada, October 1992, pp 891-893.