

OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE



Babel – Addressing the Language Deluge



L E A D I N G I N T E L L I G E N C E I N T E G R A T I O N

Mary P. Harper
Incisive Analysis Office, IARPA
Babel Proposers' Day, 20 January 2011



Agenda

9:00 – 9:15am	Introductory Remarks	Dr. Mary Harper Program Manager
9:15 – 9:30am	IARPA Overview	Dr. Peter Highnam IA Office Director
9:30 – 10:30am	Babel Overview	Dr. Mary Harper Program Manager
10:30 – 10:40am	Contract Agent Overview	Mr. Jeffrey Micher Army Research Lab
10:40 – 11:00am	Break	
11:00 – 12:00pm	Babel Questions & Answers	Dr. Mary Harper Program Manager
12:00 – 1:00pm	Lunch (Government Departs)	
1:00 – 2:30pm	Proposers' 5-minute briefings	Attendees
2:30 – 4:00pm	Posters, Proposers' Networking and Teaming Discussions	Attendees



Disclaimer

- This presentation is provided solely for information and planning purposes.
- The Proposers' Day Conference does not constitute a formal solicitation for proposals or proposal abstracts.
- Nothing said at Proposers' Day changes the requirements set forth in a BAA.
- BAA supersedes anything presented or said at the Proposers' Day.



Schedule

- Once BAA is released, questions can only be answered in writing on the program website.
- Full Proposals are due ~60 days after BAA is published.



Babel Overview

- **Goal:** Develop speech recognition capability for keyword search:
 - **Rapid development**
 - **In any language**
 - **Using limited amounts of transcribed speech**
- **Key Technical Challenges:**
 - Techniques appropriate for all languages
 - Robustness to audio collected in a variety of environments (e.g., on the street, in a car)
 - Training with limited data
 - Limited development time
 - Effective keyword search



Motivation

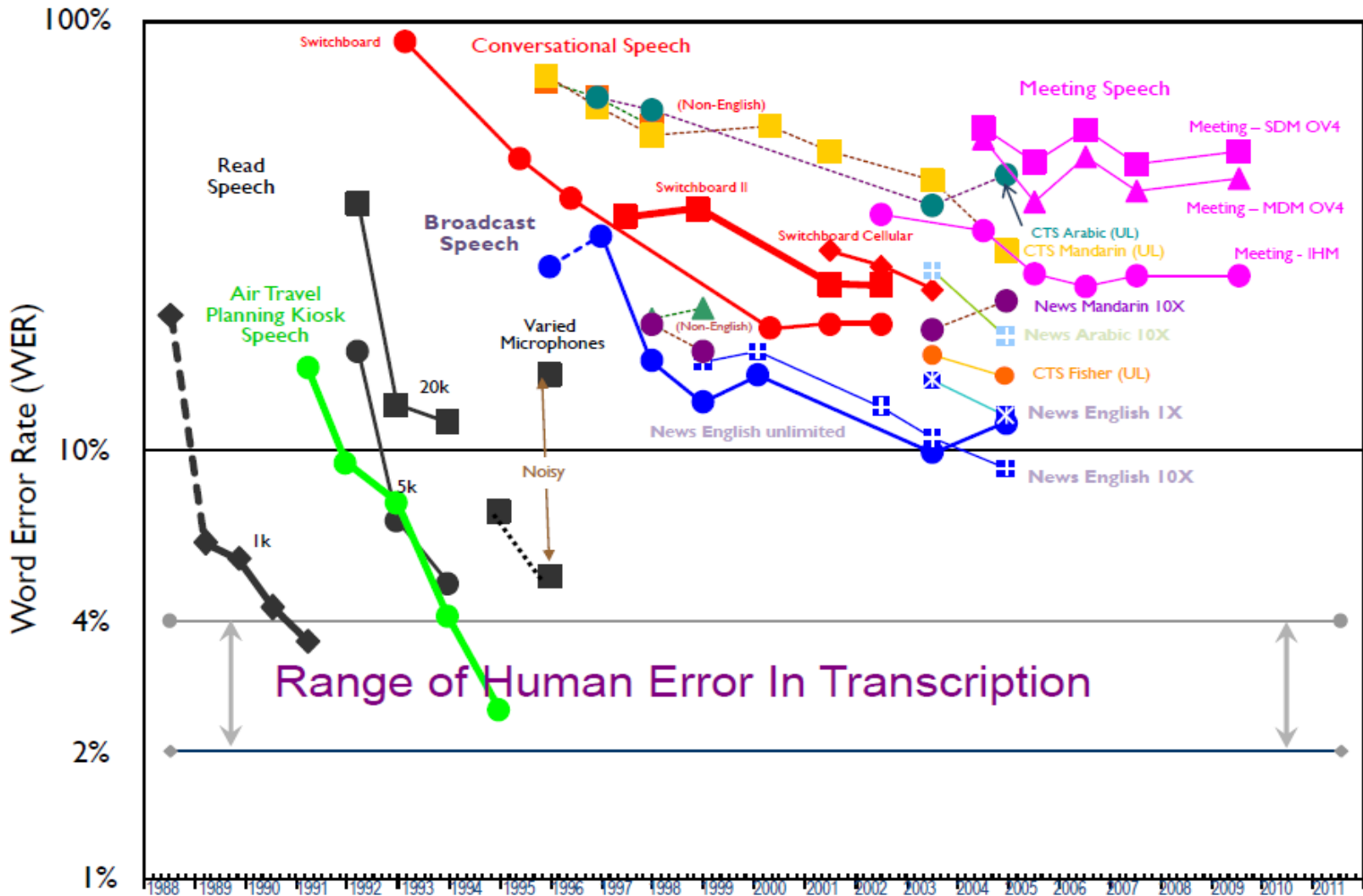
- **Speech technology research typically on:**
 - English or on languages with transcribed corpora
 - With substantial amounts of high quality audio data
- **Speech recognition technology has **limited coverage of languages****
 - 7,000 languages (from the Ethnologue)
 - of which 330 languages with more than a million speakers, but
 - speech research has focused on a handful of these languages.

Goal: effective tools for analysts to examine large amounts of noisy speech in a low-resource language under time constraints



NIST Speech Benchmark Test History until May 2009

<http://www.itl.nist.gov/iad/mig//publications/ASRhistory/index.html>



INTELLIGENCE ADVANCED RESEARCH PROJECTS ACTIVITY (IARPA)



Example Use Case

- A large volume of recorded speech is found that may contain important information, but is in a language for which there are:
 - few analysts with the ability to understand that language
 - no existing speech recognition capability to aid in analysis
- If speech recognition technology for this language could be **quickly created and deployed**, it would enable analysts to focus their efforts on the most relevant portions of the data.



Approach

- Broad language portfolio:
 - Languages from a variety of language families (e.g., Afro-Asiatic, Niger-Congo, Sino-Tibetan, Austronesian, Dravidian, Altaic)
 - Mixed language typology (i.e., with different phonotactic, morphological, syntactic characteristics)
- Researchers will:
 - work with development languages to create new methods
 - be evaluated annually on a surprise language with development time and training data size constraints
- Annual evaluation:
 - On the set of development languages and the surprise language
 - Progress will be measured using:
 - [NIST Spoken Term Detection Evaluation](http://www.itl.nist.gov/iad/mig/tests/std/2006/index.html) (see <http://www.itl.nist.gov/iad/mig/tests/std/2006/index.html>)
 - [Word Error Rate \(WER\) when appropriate for the technology](#)



Language Packs

- Each language will be provided to researchers in a “pack” that will contain speech data and language information
- Speech data may include:
 - Calls made from quiet and challenging environments including public places (bar, restaurant, shopping mall ...), street/roadside, quiet location (home, office ...), moving vehicle (in-car, train, bus ...) with handheld phone or hands-free device.
 - Mixed telephony recordings
 - Scripted speech to ensure baseline coverage of the language’s phoneme inventory
 - Conversational speech
 - Metadata for all recordings (e.g., gender, age, handset type, environment)
 - Transcription of conversational audio (the amount depends on program year)
- Language information may include:
 - Description of the language (e.g., dialect regions, phoneme set definitions)
 - Lexicon entries for words appearing in transcribed data



Program Organization

Phase 1 (2 years): Bootstrap

- Provide development packs for ~5 languages per year. The packs will cover a broad selection of languages of interest to force a general approach.
 - Y1 Packs will contain almost everything needed to build a speech recognition system except for additional text for language models.
 - Y2 Packs will contain no more than 75% transcription with correspondingly limited pronunciation lexicon
- Government evaluation will use word error if appropriate, and key word search metrics
 - Test on each development language
 - Test on a surprise language with more limited system development time and different training set sizes (e.g., 1 hour, 10 hours, 80 hours)
 - Evaluation will include data from challenging environments, including some that do not match training environments.



Program Organization

Phase 2 (2 years): Ramp up and Weaning

- Provide limited packs for ~7 additional development languages per year with no more than 50% transcribed and correspondingly limited pronunciation lexica.
- Collect more languages with less human transcription and rely on the technology developed by the performers to address the annotation gap.
- Government evaluation will use word error, if appropriate, and key word search metrics
 - Test on each development language
 - Test on a surprise language with more limited system development time and different training set sizes (e.g., 1 hour, 10 hours, 80 hours)
 - Evaluation will include data from challenging environments, including some that do not match training environments.



Program Organization

Phase 3 (1 year): Robustness and Agility

- Transition to other recording conditions (and possibly genres) to ensure robustness for the languages investigated, using
 - ~5 existing development packs will be expanded to contain small amounts of adaptation data with a different and more challenging channel
 - ~3 additional languages with speech recorded in very challenging conditions
- Government evaluation will use word error if appropriate, and key word search metrics
 - Test the ~5 development languages on the challenging adaptation channel (known language, with adaptation data)
 - Test a different development language without adaptation data on the challenging adaptation channel (known language, no language-matched adaptation data) with system development time limitations
 - Test the ~3 development languages on the matched challenging conditions (new language, hard train/test matched channel)
 - Test a surprise language with speech recorded in very challenging conditions (new language, new channel) with system development time limitations and different training set sizes (e.g., 1 hour, 10 hours, 80 hours)



Performance Goals

Year	Phase 1		Phase 2		Phase 3
	1	2	3	4	5
Transcribed %	100%	≤ 75%	≤ 50%	≤ 50%	≤ 50%
Pronunciation Lexicon	100%	≤ 75%	≤ 50%	≤ 50%	≤ 50%
Language Packs Development+Surprise	~5+1	~5+1	~7+1	~8+1	~3+1
Development Time for Surprise	1 month	3 weeks	2 weeks	1 week	1 week
Minimum NIST Actual Term Weighted Value (ATWV) ¹	0.3	0.3	0.3	0.3	0.3

NOTE: All evaluations will include data from challenging environments. There will also be alternative evaluations with different amounts of transcription.

1. See <http://www.itl.nist.gov/iad/mig/tests/std/2006/docs/std06-evalplan-v10.pdf>



Before and After Babel

	Today	After Babel
Language Coverage	Limited	Any spoken human language
Resources	100's-1000's of hours of transcribed training data	Limited amounts of transcribed data
Channel	Homogeneous	Mixed
Time to develop	Months to years	1 week



Program Roles and Responsibilities

- Performer R&D
 - In Scope:
 - Multilingual speech modeling
 - Novel use of machine learning, data resource gathering, linguistics, etc.
 - Methods to cope with limited amounts of transcription
 - Computational methods to limit running time and memory
 - Keyword search
 - Out of Scope:
 - Human user interface
 - Machine translation
- Government Support
 - Government Furnished Information (GFI):
 - Training data for diverse set of development languages
 - Development data to measure interim progress
 - Evaluation data
 - Testing and Evaluation:
 - Evaluation framework to measure performer progress on development languages
 - Evaluation framework to measure performer progress on an annual surprise language



Eligibility Information

- Other Government Agencies, Federally Funded Research and Development Centers (FFRDCs), University Affiliated Research Centers (UARCs), and any other similar type of organization that has a special relationship with the Government, that gives them access to privileged and/or proprietary information or access to Government equipment or real property, are not eligible to submit proposals under this BAA or participate as team members under proposals submitted by eligible entities.
- Non-US organizations and individuals may be able to participate.
 - Must comply with Non-Disclosure Agreements, Security Regulations, Export Control Laws, etc, as appropriate
 - Specific guidance for non-US participation will be provided in the BAA



Proposal Guidance

- Your proposal should include a full discussion of the technical approach that will be used to meet the program goals.
- Programmatic issues to be addressed in the proposal:
 - Your team's current technical capabilities
 - Key resources needed (not currently available to your team), to include capital equipment and special expertise (teaming will likely play an essential role in providing special expertise). The risk in acquiring these key resources, and mitigation strategies, should be indicated as well.
 - A teaming plan along with the roles and responsibilities of each member of the research team
 - End of phase and some intermediate milestones are set, but it is expected that other intermediate milestones that are on the critical path of the proposed approach will be offered.
 - A schedule of all milestones including a clearly charted description of the various risk mitigation strategies that will be undertaken to achieve program goals



Proposal Evaluation Criteria

- Overall Scientific and Technical Merit
- Effectiveness of Proposed Work Plan
- Relevance to IARPA Mission and Babel Program Goals
- Relevant Experience and Expertise
- Cost Realism

Evaluation criteria will appear in the BAA.



Teaming

- Because of the many challenges presented by this program, both depth and diversity will be beneficial for overcoming these challenges.
 - Throughput: Consider all that you will need to do, all the ideas you will need to test. Make sure you have:
 - Enough people and expertise to do the job
 - Sufficient resources to follow critical path while still exploring alternatives – risk mitigation
 - Completeness: teams should not lack any capability necessary for success, e.g. should not rely on enabling technology to be developed elsewhere.
 - Tightly knit teams
 - Clear, strong, management, single point of contact
 - No loose confederations
 - Each team member should be contributing significantly to the program goals. Explain why each member is important, i.e. if you didn't have them, what wouldn't get done?
 - No teaming for teaming's sake.
- Remember, you may be very accomplished, but can you do it all?



Additional Information

- Email dni-iarpa-baa-11-02@ugov.gov with additional questions.
- Babel BAA will be posted on FedBizOpps website (www.fedbizopps.gov).
- Q&As will appear after the BAA is posted. See http://www.iarpa.gov/solicitations_babel.html.



Questions?

