

A small amount of information has been removed from this report prior to publishing online.

RESEARCH

This report notes achievements and problems re our objectives of interdisciplinary training in normal speech processing by humans and machines in 4 Themes: multilinguistic phonetic detail (PD); imperfect knowledge or signal; longer speech units; exemplars and abstraction. The next Periodic Activity Report will overview all activity.

OVERVIEW

Research and deliverables are on target at individual, site and network level, given Fellows' months in post. Fellows are enthusiastic, mutually supportive, and engage readily with seniors throughout the network. Deliverables are as yet limited in number, as expected for this stage, but significant cross-fertilization in methods and theory means final output goals and overall objectives seem realistic.

QUANTITATIVE

- 16 Fellows in place, all with interdisciplinary experience and explicit training, over an average of 16 months per Fellow (range 2-20 months). Details in Annex I
- Papers: 11 in peer-reviewed journals, 10 submitted or in press, 4 in books, 37 at conferences
- 8 invitations to speak stemming directly from S2S
- 2 pieces of software delivered
- 8 pieces of software being developed that integrate knowledge from 2 or more disciplines either in a more innovative way, or with greater depth of understanding, than was achieved before S2S: Projects 2(x2), 5, 6, 7, 9, 10, 11
- ≥ 5 non-computational inter-disciplinary methods in use: Projects 1(x2), 3, 4, 8
- 9 speech corpora of L1 &/or L2 speech (7 spontaneous, 2 read; 2 of which are not yet complete) in 5 languages, that attest to interdisciplinary progress in that they are designed as a common resource to supply the quality criteria and other needs of 2 or more disciplines, and are open to a unified transcription system and to datamining using phonetically-informed automatic search
- 1 manual on corpus design and tools in preparation

QUALITATIVE

NOTABLE RESEARCH RESULTS

Theme 1: Principles for building spontaneous speech corpora in Czech, Dutch, English, Norwegian and Swedish; identification of phonetic patterns in short utterances that signal a change in topic during a conversation; frequency of reduction patterns in spontaneous Dutch; challenges to dominant views of place assimilation in French, with a key perceptual test of word recognition theories for French and English.

Theme 2: Gesture may be a more important determinant of communicative function (CF) of backchannels ("uh huh", "mm") than the auditory signal: thus ASR systems, which use the acoustic signal alone, should perhaps include multimodal signals; the largest multilingual study to date of speech perception in adverse conditions, with other preliminary perceptual data, will inform design of training regimes for L2 listeners & ASR systems for quiet vs noisy conditions.

Theme 3: Preliminary ASR software to identify syllable-like units, which improved an existing speech segmentation system based on acoustic energy and which will be combined with a new system based on sonority. Identified dependencies between properties of speech that modify pragmatic or syntactic interpretations of intonation contours, which will contribute to a more integrated theory of linguistic-communicative function.

Theme 4: Highlighted—now exploring—similarities and differences between Leuven model, MINERVA and the Perceptual Magnet Effect (PME); proof of concept for Bayesian three-way allophonic classification of vowels— allows PD that reflects the PME to be used in human perception models.

DELIVERABLES

- ASR software based on Factorial Hidden Markov Models

S2S Mid-Term Activity Report March 2009

- A new hybrid ASR system

NOVEL TECHNIQUES & TOOLS

Theme 1: In preparation: a manual about corpus design and tools to study PD; automatic alignment of acoustic signal and segments in casual speech from orthographic records. The first database to systematically address word segmentation in conversational speech, using psychological design principles to collect controlled utterances even in spontaneous speech.

Theme 2: Techniques for audio-visual annotation and search of multi-party conversations; software that makes auditory spectrograms and time-frequency masks for robust ASR.

Theme 3: Developing a decoding algorithm based on multistage graphs which takes into account potential segmentation errors; automatic techniques to characterize the dynamics of tonal contours, together with spectral information, to better identify prosodic categories in several languages.

Theme 4: Developing software to map systematic variation in subtle phonemic and morphological PD onto categorical representations, with acoustic rather than symbolic input & engineering methods informed by auditory physiology; automatic discovery of faulty orthographic transcriptions of spontaneous speech.

INTERDISCIPLINARY PROGRESS is pleasing. All projects remain interdisciplinary. P8 addressed fragmentation in prosodic research too. Most linguistics Fellows use computational or engineering methods normally unavailable to them. All ASR research includes more focussed attention to phonetic knowledge and processes; some addresses PD in depth; one spans machine and human learning. In various combinations, phoneticians work with psychologists, theoretical intonationists, conversation analysts, and computer scientists/engineers. A few Fellows successfully span 4 or 5 disciplines. Those who had no previous experience in speech work in at least 2. Several seniors began new interdisciplinary research due to S2S. Workshops fuel interaction by engaging all disciplines and scheduling social and professional networking time. We are discussing ways to deepen interdisciplinary exchange through the next period, including new grant applications.

EXPECTED BREAKTHROUGHS

Tools that allow corpus-based analyses of sound variation with different signal representations and time resolution scales, with automatic discovery of faulty orthographic transcriptions, should significantly speed up PD research, making ASR-phonetics collaboration more viable.

RECOGNITION

S2S excites international interest. Indicators include 5 invitations for keynotes etc at international events in Europe and India, raised success rates in grant applications, an Ikerbasque Research Professorship, a special issue of *Speech Communication* on Theme 2 issues, and a request to coordinate European hosting of interns from a US NSF training programme, with reciprocal hosting promised.

CITATION INDEX is uninformative within 2 years of start.

STATE-OF-THE-ART EQUIPMENT AND FACILITIES are used across the RTN. Sites collaborate in resource use: an ER organizes the corpus archive; a coordinated S2S tools archive is starting. East European sites' funded software/equipment enhances their research capability.

INDUSTRY

Two Cambridge firms joined S2S after we started. An ESR corrected labels in Toshiba's TTS database. Qualified by S2S training, she valued experience matching phonetic accuracy with the system's capability. 2 other ESRs plan industrial placements, 1 delayed for personal reasons.

INDIVIDUAL PROGRESS & ADJUSTMENTS TO PLANS

More audio-visual focus was requested and responded to by adding KTH (Sweden) to RTN activity.

The biggest problem is balancing conflicts between interdisciplinary training and contribution to general RTN aims with focussed research for a PhD/publications and hence employability. Corpus builders

S2S Mid-Term Activity Report March 2009

exemplify this tension. In response, S2S is appointing dedicated transcribers so corpora will be useful after S2S finishes, while our Fellows transcribe only what they need now.

Speech research is multidisciplinary by nature. S2S increases individuals' span and depth to avoid simplistic assumptions and methods and add insight in each area. This poses great challenges. S2S members deal well with them, but:

- 1) ER contracts are too short for our enthusiastic researchers to benefit from all interdisciplinary activity. Eg an ASR ER has no time to do planned psycholinguistic experiments; an experimental ER lacks time to contribute to new projects arising in the RTN.
- 2) The interdisciplinary skills required for some projects place great demands on ESRs.
- 3) As expected, the range of basic tutorials reduced time for discussion. This is being addressed by reviewing structure and content of future workshops.

TRAINING/TOK

QUANTITATIVE: DETAILS IN ANNEXES II, III & IV

- RECRUITMENT went mostly as planned
 - We balanced network needs and candidates' interests by interviewing candidates for more than one project if possible
 - Most Fellows were appointed within 2 months of planned start dates; all except one within 6 months
 - 98% of promised person-months will be filled; other funding will be sought for the ESR who lost 10 months due to visa delays (see Problems)
 - One ER post was split into two at the first ER's request, for family reasons. Other appointments are as planned
- Most SUPERVISION is as planned
- All ESRs plan to submit PhD theses 3-3.5 years after start of contract (excluding maternity leave)
- The 16 Fellows made 16 VISITS; 4 more are planned
- 4 visited or had online contact with the home scientific community
- All attended each RTN MEETING during their contract; 5 attended meetings before their contract began
- 11 attended TRAINING EVENTS at individual RTN sites; 11 attended external training events/conferences, 10 of whom PRESENTED PAPERS
- All presented their work at RTN meetings, covering background, progress, problems
- All ERs and 3 ESRs have TAUGHT at S2S events; 2 in a host institution
- 7 training events were organised by fellows themselves
- 6 INTERNET/COMPUTER BASED TRAINING COURSES AND SOFTWARE USER FORUMS have been developed/used

QUALITATIVE

Training and ToK are as planned or better. All Fellows meet with main SUPERVISORS, and as necessary with auxiliary ones for academic and personal mentoring. ER-ESR MENTORSHIPS exist for those who want them, but most ESRs seek help from more than one ER and from other ESRs according to topic. The frequency and value of such help is well-attested in the CDPs.

SITE TEAMS meet as needed, individuals often meeting daily or weekly. PROJECT TEAMS meet at workshops and use email/phone/Skype.

INNOVATIVE TRAINING focuses on workshops offering cutting-edge lectures, tutorials, hands-on training, discussion and social events. Fellows present work and receive feedback. Tutorials cover major areas and tools, some taught by seniors & Fellows, others by Fellows alone. ERs are keen to teach. All Fellow-led, and much senior-led, training, came from Fellows' requests and is RESPONDED TO FAST. The latest, proposed in September, was taught in January by 2 ERs, 1 ESR & 5 seniors. It was unique in breadth and depth of methods in speech research. Such courses are sometimes offered outside the RTN, but S2S projects highlight needs, and Fellows ask for training they want. They now have basic knowledge or better in most relevant areas outside their main field.

S2S Mid-Term Activity Report March 2009

The last like this is planned for April 09. Future workshop format will change to reflect Fellows' developing needs, including research productivity and employability.

COMPLEMENTARITIES between members are exploited when Fellows enrich existing work by specialist training at a second site; or visit sites with different expertise to deepen new knowledge and develop interdisciplinary research. Several projects benefit from Fellows' initiatives in exchanging ideas and practical support.

Each Fellow keeps a CDP of all ToK activity, with planned and actual outcomes, on the S2S wiki, readable by all and monitored by the Training Manager. The CDPs, together with internal PhD reports and our frequent contact at S2S training events, mean that the only other reports needed to monitor progress are those made after a secondment.

Members tell each other of OPPORTUNITIES as they arise, and post notices on the website.

Contacts within S2S enhance CAREER OPPORTUNITIES. 1 ER may continue at his host site; one has 2 interviews for academic posts and has written a grant with 3 seniors, 2 of whom he has not worked with before. No ESR is currently job-seeking, but all are aware of the potential of their wide training for their futures.

Fellows gained many COMPLEMENTARY SKILLS. All learned the host country language, and improved their English communication professionally & socially, formally & informally. All ERs and several ESRs have taught in English & extended their computational skills. Those who test human subjects learned about ethical issues. 2 ERs helped write research proposals. Those at sites hosting a workshop helped organise it. Management and personal skills are acquired by Fellows elected to the Steering Committee (SC). All Fellows may attend SC meetings; all were asked to attend when budget decisions were to be made.

There is every sign that Fellows are a mutually supportive and EFFECTIVE TEAM—sensitive to individual needs, attentive to requests for help. Individual comments indicate that the group plans ahead and takes its responsibilities seriously—and has fun. This outstanding team spirit made extra team building unnecessary.

1 ER learned to use SPECIALISED eye-tracking EQUIPMENT, all Fellows learned some new software, several did new computational modelling, and many learned phonetics & experimental methods.

The Fellows' short sketches show they are satisfied. Especially appreciated are the cross-disciplinary and personal aspects of being in S2S. Many speak of future collaboration with S2S members. One can easily imagine this group in 10-20 years forming the bedrock of research in our area.

MANAGEMENT

INTERNAL COMMUNICATION is as planned. Fellows meet often with supervisors; project teams meet at workshops and elsewhere; the coordinator liaises between the RTN and the EC; she or her assistant liaise with partners. The SC guides the RTN. Email is routinely used, with phone/Skype and the website Forum for more complex exchanges. Individual feedback is given as judged best at host sites, RTN meetings, and by phone or email.

INDUSTRIAL PARTNERS are invited to workshops. None attended yet. We will encourage them more strenuously as results appear.

The SC is the primary vehicle for RTN SELF-ASSESSMENT. It evaluates, prioritizes and delivers training and workshops, using Fellows' research papers to assess needs. It suggested and approved uses for budget surplus to achieve maximum research output with optimal cost management. The decision to include KTH members in RTN activities came from SC seniors noting evolving interests amongst ESRs. Project overlap is small and beneficial. Collaborative corpus- and tool- building are notable, overseen by the Prague ER and seniors. Benchmarks for ESRs are host-internal for PhD students. Fellow SC reps are briefed by their peers before each SC meeting.

S2S Mid-Term Activity Report March 2009

EXTERNAL COMMUNICATION is via S2S (72755 hits in 18 months), Cordis & local websites, conferences etc. A newsletter is sent to industrial partners and individuals. S2S receives emails from all over the world. Public outreach will grow as S2S develops.

THE RECRUITMENT STRATEGY was effective re open competition and equal opportunities. Most advertising was online. It achieved good international coverage: enquiries from 52 countries, about half from the EC. Interdisciplinary panels shortlist and interview following SC Guidelines. 16 interviews at the kick-off meeting led to 6 appointments. Now, interviews are mainly by Skype, but on-site for final candidates if possible. Our 16 Fellows (13 EC) are from 12 countries. Third Country appointments = 25% people, 15% person-months. Women = 44% people, 45% person-months.

MANAGEMENT DATABASES on the WIKI enable rapid access to specific information eg re ESRs, Projects, corpora. Website provider difficulties curtailed our plans but a new website, managed at Cambridge, is in place and will be finished after MTR reporting.

A CONSORTIUM AGREEMENT addresses IP. No output is commercialised yet. 2 deliverables were released under GNU license.