SHORT TERM SCIENTIFIC MISSION (STSM)
SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number:
STSM title:
STSM start and end date: 07/06/2021 to 20/06/2021
Grantee name: Emer Gilmartin

PURPOSE OF THE STSM:

Successful modelling of human conversation and production of realistic human-machine interaction depends hugely on understanding of what really happens in human talk – how various signals, including lexical, prododic and temporal, combine to maintain successful spoken interaction.
While synthesis and recognition have made great advances in recent years, there is still incomplete knowledge and understanding, particularly among technologists, of factors such as breathing and pausing and their contribution to conversation.

Stockholm has a concentration of speech technologists and linguists across a number of institutions carrying out groundbreaking research on breath and the structure of conversation (Stockholm University), synthesis of disfluencies and pausing in natural conversational speech (KTH), and the use of spoken dialog technology in education (FurHat Robotics, KTH) The STSM allowed me to travel to Stockholm for a period of at least two weeks to fulfill a number of goals.

a. Learn about breath in conversation, particularly in relation to modelling and synthesising naturalistic human talk and dialog.
b. Get hands on demonstrations and practice in measurement of breath during talk, and how corpora of natural conversation are set up and processed at SU's recording facilities.
c. (COVID permitting) Make pilot recordings of learner/native conversation with breath measurements.
d. Learn about synthesis of naturalistic conversational speech, and explore the possibilities of using this in language learning applications.
e. Explore how spoken dialog technology can be used in tutoring applications in various domains with migrant learners.
f. Exchange knowledge and experience on natural dyadic and multiparty conversation, contributing to technical understanding of the phenomena to be modelled, and gaining experience and understanding of the techniques used in the study and modelling of natural talk.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

During the two week stay, I was able to meet a number of researchers, while observing relevant social distancing practices.

At SU, I consulted with Professors Mattias Heldner and Marcin Wlodarczak on several occasions. We exchanged ideas and experience on work on the structure of conversation, and how breath contributes to this. I learned about their custom equipment for measuring respiration during speech, and how the resulting data are

processed and exploited. We also discussed the applications of this research area to more efficient artificial dialog technology.

Unfortunately, the COVID situation meant we could not use the unique facilities at SU to make recordings, azs this would have entailed bringing people together in a small recording studio for extended periods, and also because the recordings would need to be created without participants wearing masks.

A significant proportion of the stay at SU was spent designing, preparing, and performing experiments on multiparty dialogue data held at SU, to explore speech and silence dynamics and the various types of floor state change phenonena to be found in the data. Novel methods of describing and analysing the data were created. The experiments were performed on English, Swedish, and Estonian data, and produced interesting results related to how speaker activity in multiparty casual conversation relates to whether a turn ends or is retained. Part of the time was spent writing the experiments and results up for a proposed journal paper (to be submitted later in 2021).

We also spent time exploring avenues of further work in this area, identifying potential sources of funding, and preparing applications.

I consulted with Dr Eva Szekeley (KTH), on her state of the art natural speech synthesis. We discussed how this synthesis could be used to provide truly realistic models of native speaker production in conversation.

I attended workshops and demonstrations of the FurHat dialog systems, and a local phonetics conference, Fonetik.

I met with Morgan Fredrickson (Liquid.se) and Professor Joakim Gustaffsson (KTH), who organised a full day visit to Fisksätra Community Centre (Fisksätra Folkets hus), in the most ethnically diverse area of Stockholm. At the Centre I met staff and project leaders working with education and service provision to migrants in Sweden.

Prof Jens Edlund (KTH) and I discussed human machine interaction in several domains, and particularly for social good. We discussed the organisation of a workshop on Dialog for Good in early 2022, COVID depending. This workshop will be ISCA supported and will serve as a continuation of an initial workshop organised in 2019 in Stockholm.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

The main results of the stay was the establishment of strong collaborative links for different projects. All of the planned meetings and work targets were met, apart from the recording of new data (due to COVID restrictions). A very fruitful series of experiments were performed instead of the creation of live recordings.

The concrete short term scientific result was the performance of a series of experiments on how multiparty conversation progresses in terms of timing of speech and silence. These experiments and the underlying methodology are novel in the field, and the results are being written up for a submission to the Journal of Phonetics later this year.

We are also actively pursuing funding for a multi-year international project on structure of multiparty conversation, and are continuing with our experimental work. We have We hope to source enough funding to employ a PhD student to help in this work.

Consultations with Eva Szekeley and Jens Edlund have greatly increased my understanding of state of the art synthesis and, combined with an understanding of breath in conversation, greatly aid my future work in researching and building novel interactive language learning applications.

The meetings with Morgan Fredrickson (Liquid.se) and Professor Joakim Gustaffsson (KTH), and Fisksätra Community Centre (Fisksätra Folkets hus) facilitated the exchange of best practice and recent experiences in migrant integration services in Ireland and Sweden. The meetings creation of specifications for a joint project to integrate online migrant langauge and integration training with Augmented Reality and dialog technology, and plan to submit project proposals this year for a project with migrant teens in Sweden and Ireland.

FUTURE COLLABORATIONS (if applicable)

The STSM has led to a strengthening of the relationship between Emer Gilmartin and dialog researchers in Stockholm University and KTH, and has also resulted in the formation of links to Swedish agencies working in the migrant language learning and integration sector. There are a number of future collaborations arising from the Scientific Mission.

1. Emer Gilmartin, Marcin Włodarczak and Mattias Heldner are collaborating on a journal paper on dialog processes (floor state transitions)
2. We are also applying for funding in Sweden, Ireland, and through the EU for a dialog based project which will employ a PhD student and support the Pi's engagement in continuing the work.
3. I will also continue to have access to their wealth of experience and knowledge on features affecting conversational structure, particularly those related to voice quality.
4. Emer Gilmartin (and ListenHere) will collaborate with Morgan Fredrickson and Fisksätra Community Centre (Fisksätra Folkets hus) and Joakim Gustafsson to create pilot online language and integration modules using speech technology and AR, sharing the experience and ideas of Swedish and Irish practitioners in the area. This will lead to joint applications for EU and EC funding for transnational projects in this pressing area of social need.
5. The conversations with Jens Edlund and Eva Szeley will lead to further collaboration on the speech technology side, particularly around the application of realistic speech patterns in dialog to state of the art synthesis and its use in language learning activities and dialog training for migrants in Ireland and Sweden.