

FASTNET

Focus on Actions in Social Talk: Network-Enabling Technology

Technical Summary (250 words)

Networking is an essential component of human interaction, and the content of a spoken conversation has as much to do with social bonding as with the transfer of propositional meaning. Recognising that social interactions are the essential components of vocal communication, and that actions, rather than words, are the prime units to be processed in a discourse, we aim for a paradigm shift in the way computers process speech, to incorporate speaking-style information alongside message content to provide a richer expression (or understanding) of an utterance.

This research addresses the question of how speech technology should be produced which is capable of processing not only the lexical content of a spoken utterance, but also its underlying communicative intentions. People intuitively interpret the nonverbal information in speech and tone-of-voice to aid in the interpretation of each utterance in context. We aim to produce a machine that can be programmed to perform similar interpretation of speech utterances.

The research will generalise and extend previous findings from another language (Japanese) using speech data of Irish and Irish-English. The technical goal of this research is to produce speech technology specifically adapted to interactive or conversational speaking styles that will enable a friendlier and more efficient speech interface for public services, commerce, and entertainment. The academic goal of the research is to model this parallel channel of spoken communication, verifying its universality in human dialogue, but also illuminating the extent it may take on language/culture specific forms.

Lay Summary (150 words)

Current speech technology is based on text. People don't speak text, so there is often a mismatch between the expectations of the system and the performance of its users. Talk in social interaction involves the exchange of propositional content (which *can* be well expressed through text) but it also involves social networking and the expression of interpersonal relationships, as well as displays of emotion, affect, interest, etc. This research will produce a device for processing human speech, as part of a robot, an information-providing service, a translation service, or an entertainment system, which is able to process not just the text of that speech, but also able to interpret the intentions, or acts, of the speaker. It is not enough just to know what a person is *saying*; this research will help enable a machine to know what that person is *doing* with each utterance in a discourse.

Value to Ireland

The Lead Applicant has played a key role in the development of speech synthesis technology. His previous work resulted in a speech synthesis method that has now become the world standard. He was invited to join Trinity as a Stokes Research Professor in order to work more closely with the two Co-Applicants to improve signal processing of voice-quality for advanced speech synthesis. The Co-Applicants are world leaders in voice analysis research. By joining forces we envisage a new synergy which will result in a new paradigm shift in speech processing and synthesis technology.

Before coming to Ireland, Campbell worked in Japan at the Advanced Telecommunications Research Institute developing technology for conversational speech interfaces and translation devices. As part of that recent work, he collected a 5-billion-yen corpus of spoken Japanese, and developed sensor technology for multimodal conversation data capture and processing. The extension of this work forms the starting point of the present proposal. The work will provide advanced modules for the use of Irish and Irish English in a wide range of speech-related commercial applications and will boost the development of their speech technology. Whereas this is principally fundamental and basic research, it is closely linked to advanced applications and several invention disclosures, patents, and spin-off applications can be envisaged. Overseas applications of the technology will bring in licensing fees and royalties.

The Lead Applicant served as an external collaborator in the SFI-funded CNGL CSET and was committed in that context to providing advanced modules for the speech-based translation services being jointly developed by TCD, DCU, and UCD. The proposed research will provide further technology and data which is directly relevant to that work and will dovetail closely with the Co-Applicants' contributions to the same project.

The proposed research will of course provide employment and both directly and indirectly encourage young researchers in Ireland to continue furthering their education. It will employ fresh PhD graduates as post-doctoral researchers, bringing international expertise to Ireland as well as keeping Irish talent employed in the home country, and it will help towards doubling the number of PhD students, a stated goal of the Strategy for Science, Technology and Innovation, significantly increasing the number of people with advanced qualifications in science and engineering.

While the above is perhaps true of many research projects started at Trinity, the proposed research will also contribute tangibly to the future of Ireland by producing patentable technology that is applicable throughout the world and by firmly locating Irish as one of the few languages that has its own speech technology functioning at the level of the ordinary person in the street. By bringing in top-quality researchers from abroad, and training Irish graduates, this research project will boost the countrys international competitiveness with respect to speech technology and processing.

By continuing the innovative work started in Japan at ATR (the Advanced Telecommunications Research Institute International) and at NiCT (the government-run National Institute for Information and Communication Technology), Trinity will take the lead in producing modules for advanced multilingual speech technology, providing technology for devices capable of interacting with non-professionals by providing a conversational interface for the exchange of public-use and commercial information in an everyday manner.

By working closely with collaborating researchers across Dublin from UCD, from Queens in the North, and from labs in Europe and Japan, the proposed research will help to situate Ireland at the forefront of international research by encouraging visits and by the exchange of tools and methodologies, as well as by more widely advertising our results at international conferences and academic journals.

In addition to the commercial applications mentioned above, there is also a small but necessary market for teaching Irish as a world language, and the proposed research will immediately be tested in that context, providing advanced resources, both materials and technologies, for the teaching of Irish. Having a state-of-the-art conversational-speech corpus of Irish, and speech technologies tuned especially for Irish will provide the people of Ireland with necessary resources for inclusion as an equal member in the EU community.

We anticipate joining in with wider EU-funded research projects as we gain impetus locally, since there is a growing interest in this aspect of the technology in which we already have a world lead. By implementing our results in animations, robots, and information services of other countries, we will take a leading role in spreading the sound of Irish voices across the world.



Fastnet@Trinity: *An Charraig Aonair* is a symbol of Ireland, standing isolated off its coast in much the same way that Ireland itself stands isolated from mainland Europe, but serving as its westernmost component. This project, codenamed *FASTNET*, will burn a light from these shores that will be seen across the world. It will take the lead in introducing a paradigm shift in speech processing that will help ordinary people to make use of advanced technology in a simple and natural way.