

Language Resources of the Future

(a speech-based position paper)

Nick Campbell
Trinity College Dublin
The University of Dublin, Ireland
nick@tcd.ie

Speech Technology

Current speech technology is well capable of processing the mappings between speech and text, but the challenge at the current time is to model the information flow in *interactive* speech, where propositional content is conveyed along with discourse controls to simultaneously signal the speaker's cognitive attentional, intentional, and emotional states. Current applications of speech technology include general information services, customer-care, robotics, games, and interactive media content, often making use of graphical interfaces which include an avatar, embodied conversational agent, or talking heads.

Accordingly, the corpora that need to be collected for future speech technology research must be multi-modal, multicultural, and multilingual, incorporating material for the research of interpersonal communication strategies, and speech-related bodily movements, as well as the characteristics of interactive speech itself.

Speech Corpora

There is already considerable interest in the collection, annotation, and modelling of multimodal corpora, as evidenced by the number of papers reporting such work at recent Interspeech and LREC conferences, and at the special sessions of each. However, there is still little understanding of the best ways to gather representative samples of interactive speech, and of how to control the necessary variation and range of expressivity without rendering the resulting conversational interaction unnatural or artificial in any way.

The corpora of speech samples collected in the past have ranged from read speech to spontaneous, from isolated words and numbers to sentences and paragraphs, to whole stories and live telephone conversations and meetings. The focus in these collections has been on the content of the speech material rather than on the style(s) of interaction, and future collections might better concentrate on controlling the interpersonal aspects of the speech rather than on the subject matter or content of the speech itself.

Semantic Annotation

The processing of 'meaning' in speech remains a significant challenge. An ISO standard currently being considered for dialogue act annotation (ISO/TC37/SC 4 N442 rev05) proposes eleven (11) classes of discourse act, with "task-related" as the first. Task-related discourse is lexically very complex and will continue to pose a problem for speech technology, but the remaining ten (10) major categories of spoken interaction all feature limited lexical content and are characterised by very complex prosodic variability. Perhaps the most immediate task with respect to data collection for the short-term future of speech processing research is therefore how to address the problem of the sparse coverage of these ten discourse categories.

Multiculturality

Multicultural aspects of speech data collection should also be better covered in the future. It is an important concern for resource collection that the technology should serve all aspects of society, with young and old, technical and professional, educated and naive

users all being allowed equal access. However, it is not at all certain that we as a discipline have an adequate knowledge of the different needs of, or even the identity of, these different classes of society. Thus, balancing the sociological aspects of data collection will become a major concern in the near future.

Multilinguality

Universal Linguistic Rights require the provision of language services for all people in their own mother tongue. The provision of governmental, commercial, and individual services in each person's own language is a right that our technology can facilitate. However, the collection of representative resources for the less-well represented languages of the world remains a major challenge. Tools and specifications (*'blarks'*) exist for such collections, but allocating funding to cover the less-well represented languages of the world must remain a high priority.

Multimodality

Whereas the focus of speech corpora to date has been understandably on the speech itself there is a growing awareness of the multimodal nature of interactive conversational speech. Consequently, the use of video as well as audio in the collection of speech data is now becoming standard. Alongside conventional video, the use of 360-degree video and motion capture equipment is becoming common, and the amount of data in the corpus is growing at ever-increasing rates. This poses difficulties both for the annotation and storage of the material as well as for its distribution, particularly when data collections span several days or months.

Crowd-sourcing

With the growth of the internet we find an increasingly social aspect to this medium and the emergence of crowd sourcing for a variety of tasks. The annotation of speech data in this way will provide a rapid and reliable source of knowledge incorporating the views of a wide range of people from across the world, but provision of easy access to such large amounts of material remains a challenge.

Distribution

Distribution of resources will continue to remain difficult and become increasingly more so as the volume of data increases and includes multi-source video and biometric information streams. Proper structuring of metadata will allow fast access to specific subsets of the data, and by selectively varying the compression rates and podcasting the material, wider and more rapid methods of distribution should be achieved.

Rights Management

The collection of natural unstructured conversational speech material necessarily concerns the possibility of infringing personal rights when distributing such materials. The legal aspects of ownership of speech material, and the inadvertent release of personal information pose important problems that remain to be addressed.

Quantal nature of Expressivity

Speech is not a simple or complete one-dimensional information source but functions to signal multiple aspects of underspecified propositional and interpersonal communication simultaneously. The physical sciences have mastered the arts of modelling complex quantal phenomena by addressing such uncertainty without requiring any one specific interpretation to be dominant in absolute terms; we might benefit from adopting a similar approach to the processing of meaning and the interpretation of spoken communication.