Minutes and actions points from EPSRC Project Launch meeting, Stirling, 2 Oct 2015

Present: Andrew, Roger, Jon, Ricard, Amir

Summary

- Morning: mainly Introductions, presentations and broad ranging discussions. Andrew presented Stirling's AV enhancement framework; Roger Watts demo-ed his model of visual cortex processing (use of the real-time "bar codes" model of human facial feature processing); Jon briefly presented his model driven analysis-synthesis enhancement (demos at: http://staffwww.dcs.shef.ac.uk/people/J.Barker/synth.html). A discussion on corpus collection also took place.

- Afternoon: three key points raised, feature selection looking at visual to acoustic mapping. There was some discussion about the interpretation. The full process of audiovisual filtering will need agreement over the features and processing methods to be employed/developed, and this needs to be determined in an organised manner. The limitations of existing datasets were also discussed. Task definition was also generally discussed, focusing on the need to define the task domain before continuing research. At the moment, the proposal presents a number of ambitious ideas (such as integrating, possibly at multiple (fine and coarse) scales, Stirling's and Sheffield's contrasting yet complementary approaches to speech enhancement: noise filtering vs. gap filling/re-synthesis etc.), but there is a lot of room for clarification and further emphasis. Thirdly, the corpus was discussed in more depth.

Key Points

1. Barcode models

- Discussion of computational human vision model by Roger, real time barcodes modelling faces and extracting visual features. Potential for further use and development, but will need to investigate practical feature extraction and then formatting for future work. Further meetings will be needed for this.

Actions:

-Andrew to arrange to meet Roger to better understand his model and obtain matlab demo code for possible reuse/integration into our future system (through further development/adaptation work as appropriate).

-Amir to look into setting up joint-honours/MSc projects with Roger for further development of his vision model.

2. Project research contributors

- Discussion of Sheffield RA recruitment situation:

Actions: Need to firm up Ricard's start date. Sheffield PhD student being recruited for Oct 2016.

Jon to confirm to Amir after discussing with his Sheffield colleagues

3. Feature selection

- Andrew presented some interesting preliminary work looking at visual to acoustic mapping. There was some discussion about the interpretation, i.e. in real use-case, we will need visual signals that are most informative about the missing audio information when conditioned on the present audio. These may not be the same as visual features that are best able to fully reconstruct the audio. The full process of audiovisual filtering will need agreement over the features and the processing method, and this needs to be determined in an organised manner. It was emphasised that this was just preliminary work to establish a baseline, rather than representing the final approach, as the research in this area is quite limited, with more focus on speech recognition, and limited use of both linear and non-linear modelling approaches.

Actions:

- Andrew to finalise preliminary work (as part of summary progress/review report to follow –see below) to use as a possible baseline for future project research.

4. Corpus Discussion

- Corpus discussion, issue raised that the project schedule needs to be adjusted to better take account of the corpus collection process. Some discussion of content and recording techniques, Agreed with Amir's suggestion that this collection needs to be arranged earlier than originally scheduled in the Workplan.

-Roger discussed the facilities and subjects available via the psychology department, Andrew discussed the previous corpus collection that had taken place in Stirling. Some discussion of different approaches to data collection, including conversational and reading tasks, free speech, and sentence reading.

- Roger discussed their use of a mirrored setup to enable the recording of conversational speech with two people. Amir suggested to discuss/explore benefits of such recordings with Bill/IHR and Peter/Phonak.

- Jon and Ricard discussed the facilities available at Sheffield, as they have an AV recording booth there that can be used.

- Amir pointed out that the gannt chart has data collection in year 2, which will be too late, as we need the data for model development. Hence need to start planning sooner. Data collection seems like it will be relatively straightforward, but the biggest issue is designing the corpus and deciding what content/format to include.

- It was noted that AV data are needed for two separate purposes: training models (e..g. V->A mappings etc) and for evaluation. How best to design a corpus that's suitable for both purposes?

Actions:

- Andrew to revise Gantt chart to reflect this (include in summary report to follow - see below)

- Andrew and Ricardo to follow up (by investigating appropriateness of facilities for meeting our AV corpus requirements/specifications, including through meeting/discussions with Bill and Peter - see below)

5. Task definition

-some general discussion about the need to define the domain of the task we are tackling before doing further serious work. At the moment, the proposal presents an idea, but there is a lot of room for clarification and further emphasis. We need to make this a priority to ensure the focus is kept on the project.

- Danger of setting sights too high and failing, (could reduce chances of future funding), or setting them too low and results not being relevant.

- Degree of visual noise - We can target scenarios at different positions on this axes but Jon's view is that we should keep the visual signal 'clean' but realistic. Life is full of visual occlusion, variable lighting etc, but there are some acoustically noisy scenarios when the visual data is relatively clean, e.g. a two party face-to-face conversation in a noisy cafe and where the participants know that lip-reading is being used.

- Andrew and Amir of the view that based on preliminary work, it would not be impossible to construct an approach to take account of brief periods of occlusion, using fuzzy logic techniques and detectors (including audio and visual spike based) as presented in preliminary research, and so handling visual noise could potentially be easily done and demonstrate greater feasibility.

- Ultimately this depends on the exact task to be solved.

- Intelligibility versus quality -- It was noted that these two objectives are often in opposition, e.g. increasing intelligibility may require distorting the speech . Andrew noted that quality is important as users won't tolerate a device if it is not high. Research has shown that those with hearing aids do not like poor quality signals, even if intelligibility is improved.

- If we can improve intelligibility without damaging quality then great, but if not then how do we weight these factors? Do different scenarios place a different premium on intelligibility versus quality? Can we imagine a control setting where the processing will be tuned to favour one or the other.

- Amir pointed to FAAF (Four Alternative Auditory Feature) intelligibility tests he had used in the past (supplied by Nottingham IHR), and highlighted need to review the literature. What materials are used to judge audio-only hearing aid processing and enhancement systems? Is there suitable AV equivalents that we might be able to obtain?   Further input needed from specialists.

- Generally, with hearing aids, SNR gain is often advertised as the "performance", whereas with other speech filtering approaches, intelligibility is considered more important. Amir iterated need to consult with IHR and Phonak.

- As we are considering this from the perspective of hearing aid technology down the line, this is key to the evaluation, and so needs to be thought about carefully

- Intelligibility and quality assessment probably require different types of speech material: i.e. intelligibility is hard to measure for conversational speech and typically uses isolated words, or structured sentences (e.g. FAAF tests). Quality might be best measured using natural conversational speech.

Actions:

- Amir emphasized need to contact other partners on this, particularly clinical (MRC IHR) partner, Bill Whitmer, (and Phonak's Peter) to enquire about future intelligibility testing requirements for our prototype, and also get expert advice on any trade-off between intelligibility vs quality we may need to consider/specify, when planning future AV hearing aid prototype design/development and (clinical) evaluation. Amir and Andrew to follow-up (including through circulation of draft summary report to be put together by Andrew, for Bill's and Peter's comments/input)

- Link above (in summary report) to trade-off between intelligibility vs quality discussed earlier

Final Points

- start preparing a report summarising data requirements for hearing aid assessment, importantly, by further reviewing state-of-the-art and consulting Bill and also Peter as noted above: ACTION: Andrew

- summary report needs to compare new AV Corpus design and development requirements against currently available AV data sets. This report will be used to inform the design of our own corpus collection and design of speech intelligibility and quality assessment/evaluation experiments - in partnership with Bill & Peter. ACTION: Andrew

- finalize draft agreed minutes of inaugural meeting for circulation to all, including Bill and Peter (by end of October) ACTION: Amir

- Contact Gerasimos Potamianos regarding possible availability of IBM's AV corpora. ACTION: Jon

- Contact Edinburgh to assess appropriateness/availability their MC-WSJ AV corpora. ACTION: Jon

- Draft of summary report to be prepared and circulated to all partners (during November), and Bill and Peter to be invited to next all-hands project meeting at Stirling, possible towards end of November/start of December ACTION: Andrew, Amir and Jon

- In line with original workplan, Andrew to arrange travel to Sheffield (during early November) to meet Ricard (and Jon), to better understand their work and explore links (both those discussed at this meeting and those proposed in the original proposal) and include relevant (action/discussion) points in forthcoming summary report (due in November), for further discussion at next all-hands project meeting, during late Nov/early Dec. ACTION: Andrew

- Amir and Jon to draft minutes (and action points) at future project meetings and reconcile/finalize/circulate within a week of meeting date.