SSPNet
Social Signal Processing Network
http://www.sspnet.eu/

Funded under the 7th FP (Seventh Framework Programme)
Theme ICT-2007.2.2
[Network of Excellence]

D10.2: Progress report, compendium of work done
during months 1-12

Due date: 31/01/2010  Submission date: 29/01/2010
Project start date: 01/02/2009  Duration: 60 months
WP Manager: Steve Renals  Revision: 1

Author(s): R. op den Akker, J. Carletta, M. Mehu, C. Pelachaud, O. Türk, M.F. Valstar, T. Wilson

<table>
<thead>
<tr>
<th>Dissemination Level</th>
<th>PU</th>
<th>RE</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public</td>
<td>Restricted to a group specified by the consortium (includes Commission Services)</td>
<td>Confidential, only for members of the consortium (includes Commission Services)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
D10.2: Progress report, compendium of work done during months 1-12

Abstract:
This document reports progress in WP10 during the first year of SSPNET.
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2 Integration efforts</td>
<td>7</td>
</tr>
<tr>
<td>3 List of achievements</td>
<td>8</td>
</tr>
<tr>
<td>4 List of future achievements</td>
<td>24</td>
</tr>
</tbody>
</table>
1 Introduction

Within SSPNet, WP10 concerns the analysis and interpretation of social signals during face to face meetings. The objective of WP10 is to model, detect and interpret automatically social signals displayed during these meetings. The three main goals identified are to 1) automatically model and infer social context information (e.g. the roles of the participants) in multiparty meetings, 2) incorporate social signals (e.g. focus of attention, facial expression) in multimodal recognition and synthesis, and 3) demonstrate the results in two applications: the analysis of group interactions involving several people, and the synthesis of realistic groups of virtual characters.

The real work this year has been to lay the groundwork for collaboration for this new and challenging area of research. Our main activities have been:

- development of a shared understanding of the AMI Meeting Corpus. This sounds trivial, but the network comes from a wide range of disciplines and levels of technical expertise — just getting to grips with each other’s work and the possibilities for this new kind of data take time.
- the changes to Greta that make it possible to control it at a finer grained level. This raises the opportunity of using an embodied conversational agent to study turn-taking and other social aspects of meetings, which was not possible before.
- annotations that will let us move forward from the relatively easy task of figuring out the assigned task role in a design team (project manager, industrial designer, user interface designer, marketing expert) to the more useful one of what social roles people play in a meeting.

2 Integration efforts

At present, the analysis and synthesis of behaviour in face-to-face meetings is stuck at a rudimentary level. Computationalists have so far dealt with exaggerated data and stylyzed behaviours. Analysts describe instances of behaviours that they see, and build some generalizations, but not yet in a way easily amenable to computation. We can only make advances by fostering new interdisciplinary links.

Our strategy has been to encourage partners from all the component disciplines to look at the same two data sets, the AMI Meeting Corpus and the Canal 9 political Debates. Although these datasets have their limitations like any other, our past experience is that bringing people to the same data from different perspectives helps people to understand each other’s research and develop new research questions and strategies that would not arise independently. As one would expect within a network, our thematic workpackages are highly interconnected. In WP8, the object of study is political debates. Debates and meetings are amenable to many of the same kinds of processing and analysis, but the differences between debates and meetings makes in an open research question whether strategies developed on one will transfer well to the other. In general, there has been more work previously on visual processing in debates and on language in meetings, and many network members have been more familiar with the debate data.

The application of the same approaches over the two above datasets is the basis of the interconnection between WP8 and WP10. The data used in the two WPs is close enough
for testing generalizability but keeps the community from falling into the trap of relying too heavily on the idiosyncracies of a single genre. Infrastructural developments in WP8, such as new processing techniques, are relevant to WP10 because the same techniques can be tried on meeting data, either by network members or by the wider community. The same applies to approaches developed for the meeting data that can be beneficial to the analysis of political debates. WP9, about politeness, addresses behaviours that cut across a wide range of genres, including that of WP10. Where WP9 work is about face-to-face interaction, it is relevant to WP10, even it is about pairs of conversants — since small group meetings can be viewed as conversations with a changing pair of speakers and set of overhearers. As a consequence of such differences, the AMI corpus is typically more apt for the analysis of floor and turn-taking, as well as for the analysis of emergent role (different from those stereotyped or institutionalized of Canal 9); but both AMI and Canal 9 are amenable for modelling (dis)agreement and dominance.

As we have identified pairwise and small group interactions that make the most sense to the researchers involved, coordination has shifted to ensuring that these interactions have the infrastructure and information they need to make progress. Because of this strategy, much of the communication in the workpackage has been by email and smaller phone conversation. We met face-to-face on 13 Feb and 3 Dec 09, and held larger telephone conferences on 10 Mar, 1 Apr, and 7 Jul, in which we concentrated on what the various disciplines found of interest in the corpus data.

Deliverable WP6.2 reports about the exchanges that took place between partners involved in WP10 and have been aimed at activities relevant WP10 framework:

1. University of Edinburgh - CNRS
2. Idiap - University of Twente
3. Idiap - University of Roma Tre

Furthermore, the WP10 members had the opportunity of meeting at the workshops organized by the SSPNet, namely the Workshop on Social Signal Processing (Sept. 2009) and the Workshop on Foundations of Social Signals (Dec. 2009).

Our records and plans are kept on a workpackage intranet in order to give us better group memory and encourage new participation.

3 List of achievements

1. Development and release of a tool for identifying subjective utterances in multiparty conversation

   **Description:** The tool for identifying subjective utterances in multiparty conversation was developed using the subjectivity annotations currently available on the AMI Meeting Corpus. The tool combines models trained on words, characters, and phonemes to classify utterances as subjective or objective.

   **Tangible outcome:** Tool for identifying subjective utterances in multiparty conversation
Partners involved: University of Edinburgh

Dissemination: A link to tool for identifying subjective utterances is available on the SSPNet portal. Automatic subjective utterance annotations produced on the AMI Meeting data are also available.

Related Publication(s): N/A

Links to other thematic WPs: N/A

2. Investigation of automatic agreement detection in multiparty conversation

Description: We investigated the automatic detection of agreements and disagreements in multiparty conversation using a variety of lexical, structural, and prosodic features and two different machine learning approaches. We compared the performance of a decision-tree learner and a graphical model (conditional random fields), and we also explored the use of hand-crafted, high-precision rules in a pre-processing step as a means of reducing the effects of the data imbalance.

Tangible outcome: Paper.

Partners involved: University of Edinburgh, DFKI

Dissemination: Talks were given at ICMI/MLMI and the SSPNet Rome workshop on the agreement detection work.


Links to other thematic WPs: N/A

3. Social role annotations on the AMI Meeting Corpus

Description: Using the annotation manual from FBK, we are annotating social roles in the AMI Meeting Corpus. Two pilot annotation studies will be used to evaluate the applicability of the social roles annotation scheme for the AMI data. The current goal is to annotate a portion of the AMI meeting data with the social roles of the participants by the end of February.

Tangible outcome: Adapted social role annotation manual in English. Social role annotations on the AMI Meeting Corpus, expected end Feb 2010.
Partners involved: Edinburgh, IDIAP, as well as a developing collaboration with FBK

Dissemination: N/A

Related Publication(s): N/A

Links to other thematic WPs: N/A

4. Development of an approach for automatic role recognition in conversations

Description: We have developed an automatic approach for the recognition of roles in broadcast data (news and talk-shows) and in meetings. In broadcast data, the roles correspond to the functions that people fulfill in the structure of a given program (e.g., 'Anchorman', 'Guest', 'Headline Person', etc.). In meetings, roles correspond to the position of meeting participants in an ideal company in the framework of which the meetings are held (e.g., 'Project Manager', 'Marketing Expert', etc.).

The approach includes three major steps:
- Segmentation of the data into single speaker intervals, i.e. time segments where only one person talks or holds the floor. The result is the turn-taking, i.e. the sequence of turns composing the conversation portrayed in the data.
- Extraction of a social network from the turn-taking. People that talk in the same time intervals are supposed to interact and this information is sufficient to extract a social network. The network is used to represent each interaction participant with a feature vector.
- Role assignment. This step assigns a role to each interaction participant using machine learning and pattern recognition (in particular, Bayesian classifiers and Hidden Markov Models).

The results show that roles are recognized with an accuracy (percentage of time correctly labeled in terms of role) of around 80% in the case of broadcast data, and of 45% in the case of meeting recordings.

Tangible outcome: - A package for analysis and representation of turn-taking - Publications (see below)

Partners involved: Idiap Research Institute

Dissemination: - The publications are available on the SSPNet Portal

Links to other thematic WPs: This work is relevant to WP8 because roles are a key aspect of political debates as well.

5. Identifying social factors in floor management

Description: We try to disentangle the social factors that are involved in turn and floor management in conversations. This sheds a light on different aspects that go into politeness (rules of etiquette, personality, role, dominance, rudeness, rapport, etc.). We studied the literature on floor with this perspective in mind. The analysis will be used in the synthesis of behaviours for conversational agents with different politeness strategies.

Tangible outcome: Presentation at workshop (see Virtual Learning Center). The tangible outcome will include a review paper and an annotated bibliography on the portal.

Partners involved: University of Twente

Dissemination: The work was presented at the Foundations of Social Signal Processing in Rome.

Related publications: N/A

Links to other thematic WPs: The work is related to the definition of a floor annotation schema in WP10.

6. Progressive development of a framework for describing social signals

Description: We have developed a framework that sets out in a reasonably compact way the main issues that the relevant literatures in the human sciences (mainly psychology, anthropology, and linguistics) indicate are relevant to understanding the processes of generating and interpreting indicators that may carry information about social states, relationships, intentions, etc.

Tangible outcome: N/A
**Partners involved:** QUB, CNRS, U. of Twente, DFKI

**Dissemination:** Successive versions of the talk have been presented at ACII09 (see publication), in the WP meeting in Paris (Nov 2009), and in the Workshop on Foundations of Social Signals (Rome, Dec 2009). The ACII and Rome presentations are available on video.


**Links to other thematic WPs:** The issues affect all WPs, and most partners have been involved in the discussions that have shaped the current form.

7. **Applying the general framework for describing social signals to the particular case of politeness**

**Description:** A framework has been developed which integrates diverse (and in some respects contrasting) analyses of politeness in the literature, distinguishing different types of politeness and the different levels at which it may be signalled.

**Tangible outcome:** N/A

**Partners involved:** QUB, CNRS, U. of Twente, DFKI

**Dissemination:** P Brunet presented a paper on politeness in the WP meeting in Paris (Nov 2009), and the material was included in his paper to the Workshop on Foundations of Social Signals (Rome, Dec 2009). The Rome presentation is available on video.

**Related Publication(s):** N/A

**Links to other thematic WPs:** The work on politeness is deliberately being cast in a format that is relevant to all the thematic workpackages.

8. **A survey of the nonverbal auditory and visual cues that could be present during (dis)agreement and of possible tools to detect them**
**Description:** This was the first step towards achieving our eventual objective: the automatic detection of (dis)agreement based on the presence and temporal dynamics of the relevant behavioral cues. A review of the Social Psychology literature was completed in order to survey all visual and auditory cues which could be present during (dis)agreement. Furthermore, a list of tools that could be used for the detection of these cues was compiled.

**Tangible outcome:**
- Publications (see below).
- Bibliography of 83 references
- A list of tools ready to use towards the WP8 objective, after minor adaptations.

**Partners involved:** Imperial College, University of Geneva

**Dissemination:** The corpus is available via the following URL (to be copied to SSPNet web portal as well): [http://www.doc.ic.ac.uk/~kb709/a_da_bib.htm](http://www.doc.ic.ac.uk/~kb709/a_da_bib.htm)


**Links to other thematic WPs:** Agreement and disagreement are one of the most common interactive signals and, hence, revealing agreement and disagreement is useful for analysis of any interpersonal interaction including group interactions (WP10).

**9. Collection of a bibliography on linguistic aspects of social interactions (with particular attention to agreement and disagreement)**

**Description:** We have identified the linguistic aspects most important for the analysis of political debates and, in more general terms, of conversations. We paid particular attention to (dis-)agreement, dominance, status, role, and emotions. Collected references come in particular from the following areas: conversation analysis, pragmatics, phonetics and linguistics. This work has been useful to identify expertise gaps in the consortium as well.

**Tangible outcome:** bibliography containing 150 references

**Partners involved:** Idiap Research Institute, Universita’ Roma Tre

**Dissemination:** The bibliography is available on the portal.

**Related Publication(s):** N/A
Links to other thematic WPs: Useful for group interactions (WP10).

10. Review and bibliography of floor and face

Description: We have written a review of the history of linguistic research on the concept of floor and face in conversations.

Tangible outcome: Paper in draft form

Partners involved: U. of Twente

11. An analysis and specification of overlapping speech in the AMI Meeting corpus

Description: An analysis of overlapping speech occurring in the AMI Meeting corpus was carried out. An existing overlap annotation scheme was applied to five meetings. The analysis in terms of interruption behavior showed that there is a correlation between dominance and interruption behavior, but the relation between speaker role (e.g., project manager) and interruption behavior is weak.

Tangible outcome: Overlap annotation scheme. Presentation at workshop (see Virtual Learning Center)

Partners involved: University of Twente

Dissemination: Presentation at the Foundations of Social Signals workshop in Rome 3-5 Dec 2009

Related Publication(s): N/A

Links to other thematic WPs: Interruption behavior can be evaluated in terms of politeness (WP09).

12. A Theory of Meeting Conversations and Activities

Description: We are developing a theory of meeting conversations that describes how floor and turn taking are organized. The theory is based on the notions of activity (Jens Allwood) and floor (Edelsky, Hayashi). The theory describes how the various relations between activities (plan, about, part of, among others) and how activities are organized by participants. We introduce the notion of arrangement, an agreement concerning the allocation of turns in a group activity. Activities are performed on a floor. We illustrate the multi-layered structure of activities and floor and how this influences turn-taking and floor operations. These illustrations come from a detailed floor analysis of a number AMI meetings. The theory describes how floor is established dynamically and how floor changes are negotiated by verbal and non-verbal boundary moves of those involved.
Tangible outcome: N/A

Partners involved: UT, UEDIN

13. The use of turn-taking patterns and prosodic features for the recognition of speaker roles in broadcast material

Description: Turn-taking patterns and prosodic features were used with Conditional Random Fields (CRF) to recognize speaker roles (e.g., anchorman, weather man etc.) in broadcast material. The turn-taking patterns appear to be strong predictors of speaker role. The addition of prosodic features to the turn-taking features do not significantly increase performance.

Tangible outcome: N/A

Partners involved: Idiap, University of Twente

Dissemination: Presentation at the Foundations of Social Signals workshop in Rome 3-5 Dec 2009

Related Publication(s): N/A

Links to other thematic WPs: N/A

14. Annotating Floor in Meeting Conversations

Description: In order to develop a theory of floor and face in meetings we developed a floor annotation scheme and performed an annotation exercise in a number of steps on three different AMI meetings with a number of annotators. We annotated floor status and floor operations and performed a reliability study and a detailed agreement analysis. The analysis form the flesh and bones of a multi-layered model of activity and floor (cf the work of Grosz and Sidner in discourse theory) that we describe in a forthcoming paper.

Tangible outcome: Floor state and floor operations annotation procedure. Floor annotations of three AMI meetings

Partners involved: U. of Twente

15. Bibliography collection on politeness (mostly speech related)
**Description:** We have made a preliminary literature survey (mostly speech related) targeting the following aspects: voice quality and politeness, intonation and politeness, non-verbal vocalisation and politeness, context-dependence in polite behaviour, cross-cultural and cross-language differences on politeness. We have identified measures as well as procedures to analyse prosody and voice quality on controlled data and natural speech recordings that can be useful for us to relate features with social signals of concern.

**Tangible outcome:** Bibliography containing 60 references.

**Partners involved:** DFKI

**Dissemination:** The bibliography is available on the SSPNet portal at: https://wcms.inf.ed.ac.uk/sspnet/wp-spaces/wp9-webspace/dfki-ideas-and-goals-for-wp9-preliminary

**Related Publication(s):** N/A

**Links to other thematic WPs:** WP9, WP10

16. **Discussion of issues regarding a markup language for social signal processing**

**Description:** Preliminary considerations regarding an XML-based markup language for social signal processing has been presented for discussion in the WP9 meeting in Paris on November 2009. This is a markup language view of QUB’s proposal of relevant variables. The aim is to represent variables relevant for social signal processing in a way that they can be used in technology dealing with social signals. Useful comments were received during the discussion.

**Tangible outcome:** Presentation slides discussing concepts underlying a possible XML markup language for SSP

**Partners involved:** DFKI, Queen’s University Belfast

**Dissemination:** Presentation in the WP9 meeting in Paris on November 2009.

**Related Publication(s):** N/A

**Links to other thematic WPs:** WP9

17. **Harmonics+Noise Model (HNM) baseline implementation**
**Description:** We have completed a baseline implementation of a harmonics+noise model (HNM) under the open source MARY TTS system. This parametric model supports more robust modification of prosody to enable larger amounts of prosody modifications with less reduction in naturalness and quality on the synthetic speech. This HNM model as well as speech signal modification algorithms will allow us to generate target prosody and voice quality patterns during synthesis and will be integrated with the speech synthesis engine in MARY TTS. Target prosody and voice quality patterns will be analysed/extracted from real data (see achievements 5 and 6) in the form of expert rules that relate features with social signals of concern.

**Tangible outcome:** A package for HNM speech analysis and synthesis

**Partners involved:** DFKI

**Dissemination:** Freely available on MARY TTS version 4.0: http://mary.dfki.de/Download

**Related Publication(s):** N/A

**Links to other thematic WPs:** WP9, WP10

**18. Implementation of voice quality and prosody measures relevant to social signals**

We have implemented 30 measures, reported in the literature as robust and effective for discrimination of voice quality and prosody. We have validated these measures with controlled data: the NECA database [Schröder and Grice, 2003] which is a diphone database that contains a full diphone set for each of three levels of vocal effort, for simplicity referred as "soft", "modal" and "loud"; and the Berlin database of emotional speech [Burkhardt et. al., 2005] which contains recordings of 10 speakers (actors), speaking in neutral style and 6 different emotions: angry, bored, disgust, fearful, happy and sad. Using the NECA DB, the implemented measures and a principal component analysis, we were able to confirm the perception tests reported in [Schröder and Grice, 2003]. As it is shown in the Figure 1, the first two principal components give us a clear separation of the three effort ratings, for the female speaker (lower part) and the male speaker (upper part). Regarding the emotions on the Berlin database, we were able to replicate, to some extent (similar means but larger standard deviations), some of the voice quality measures reported in the experiment of [Lugger et. al., 2006]. We have also developed a framework to analyse prosody and voice quality of social signals in the AMI Meeting corpus. The framework includes:

- Use of available AMI corpus annotations
- Use of the NXT tool for querying and extracting dialogue acts and speech segments from the corpus.
- Extraction of 30 voice quality and prosody measures, frame based and utterance based.
Figure 1: Principal component analysis results for three voice qualities in the NECA DB.
Performing principal component analysis (PCA) to search for patterns and the best discriminators among the measures.

**Tangible outcome:** A framework for analysis of voice quality and prosody of speech based on high quality free available software and tools.

**Partners involved:** DFKI, University of Edinburgh

**Dissemination:** Presentation at Rome SSPNet Workshop on foundations of Social Signals

**Related Publication(s):** N/A

**Links to other thematic WPs:** WP9, WP10

19. Analysis of voice quality and prosody of speaker roles in scenario meetings

Using the measures and framework described above, we have analysed the voice quality and prosody of four different roles: project manager (PM), user interface designer (UI), industrial designer (ID) and marketing expert (ME), in 9 IDIAP meetings of the AMI meeting corpus. We have analysed the speech of one word that appears frequently along the corpus, the word control. We used just one word in order to reduce contextual variability by holding the phonetic segmental form constant. We have found 373 segments containing the word control spoken by 36 different speakers on the four different roles. Figure 2 shows the patterns found with this analysis. The analysis has shown a tendency of the project manager to speak with a loud vocal effort and the marketing expert to speak with a soft vocal effort. The analysis has also shown that the measures that better correlate with these observations are skewness gradient, glottal opening gradient and fundamental frequency, so we can use this information in our synthesis algorithms to generate these two effects.

**Tangible outcome:** Two rules that can be applied on our speech signal modification algorithms and harmonic plus noise model to synthesise the speech of a project manager and a marketing expert according to the effects found in the model.

**Partners involved:** DFKI, University of Edinburgh

**Dissemination:** Presentation at Rome SSPNet Workshop on foundations of Social Signals

**Related Publication(s):** N/A
Figure 2: Principal component analysis results for prosody and voice quality measures of four roles in the IDIAP meetings.
20. Analysis of voice quality and prosody of dialogue acts in scenario meetings

Using the measures and framework described above, we have analysed the voice quality and prosody of the speech of four types of dialogue acts: inform, suggest, access and elicit, in 9 IDIAP meetings of the AMI meeting corpus. In this case we have analysed the speech of the whole dialogue act sentence, 14,304 speech segments in total. We have also analysed average measures per user and dialogue act type to reduce variability of the data, that is 144 averaged measures corresponding to 36 speakers (male and female) and 4 dialogue acts. Figure 3 shows the patterns found with this analysis. The analysis has shown a tendency of people to use a loud vocal effort when speaking assess dialogue acts. No particular pattern was found for the other three dialogue acts. According to the analysis the measures that better correlate with this observation are the incomplete of closure and fundamental frequency.

Tangible outcome: One rule that can be applied on our speech signal modification algorithms and harmonic plus noise model to generate a loud vocal effort when synthesising the speech of assess dialogue acts.

Partners involved: DFKI, University of Edinburgh

Dissemination: Presentation at Rome SSPNet Workshop on foundations of Social Signals

Related Publication(s): N/A

Links to other thematic WPs: WP9, WP10

References:


21. Adaptation and voice conversion for the synthesis of child speech
Figure 3: Principal component analysis results for prosody and voice quality measures of four dialogue acts in the IDIAP meetings.
**Description:** This study compares two different methodologies for producing data-driven synthesis of child speech from existing systems that have been trained on the speech of adults: HMM-based speech synthesis, and unit selection. In a subjective evaluation of the similarity of synthetic speech to natural speech from the target speaker, the HMM-based systems evaluated are generally preferred, although this is at least in part due to the higher dimensional acoustic features supported by these techniques.

**Tangible outcome:**
- Interspeech paper (see below)
- Implementation of text-to-speech synthesis for a child’s voice

**Partners involved:** University of Edinburgh (also: Inline Internet Online Dienste GmbH, Germany)


**Links to other thematic WPs:** This is closely related to WP9.

**22. Digital MEMS microphone array**

**Description:** A digital MEMS microphone array was designed, implemented and tested on a speech recognition. This is the first microphone array built using such technologies using commodity hardware. It has the potential to be much smaller, cheaper and more flexible than existing microphone arrays. It thus makes much more possible the ability to cheaply instrument environments to capture audio unobtrusively, important to capture data from social situations. We tested the array on a large vocabulary speech recognition task, comparing with a high quality array of analogue microphones. Initial results employing no adaptation show that performance using the digital array is significantly worse than the analogue device. Subsequent experiments using channel and speaker adaptation reduced this gap, from about 14% higher word error rate to 4.5% higher word error rate.

**Tangible outcome:**
- ICASSP paper (see below)
- Design and construction of a digital MEMS microphone array,
Partners involved: University of Edinburgh (also: EADS Innovation Works)

Dissemination: Demos given at Edinburgh to academic and commercial visitors.


Links to other thematic WPs: N/A

4 List of future achievements

Our achievements in the first year can largely be seen as preparation for achieving research results in the second year. Because this is a interdisciplinary network for work that is ambitious with respect to the state-of-the-art, we cannot promise specific research results. However, the following collaborations are currently in progress:

Interactivity measures. We are considering interactivity measures that show the degree to which group members share vocabulary and whether there are predictable patterns of interaction or everyone contributes freely, speaking, for instance, before and after everyone else in a meeting. In the social psychology literature, such measures have been tentatively linked to innovation and group effectiveness. The measures may be useful for user displays that intervene in the group process, or as features for learning useful ways of segmenting the meetings, for instance.

Role recognition. WP8 has developed technology for recognizing the role a participant takes in a debate, such as anchorman or chairman. We intend to adapt this kind of technology for the social role information for the AMI Corpus defined in the first year and currently being annotated. Automatic social role recognition is relatively challenging, but this makes it a useful point of contact for the contributing disciplines.

Subjectivity. We will look at the relationship between expressions of opinion during meetings, role, and how people interact with each other, in order to get a better idea of how one might transfer text opinion-mining technologies to multimodal data.

Behaviour analysis. We will explore the relationship between discourse structure, nonverbal behaviour, and participant demographics in the AMI Corpus, in order to develop a better, corpus-based account of overall group behaviour.

Turn-taking synthesis. Using the ECA infrastructure developed in the first year of the network, we intend to implement a simulation of naturalistic turn-taking in face-to-face meeting, based on turn-taking patterns derived from the AMI data.

Floor. Pre-existing accounts of who has the floor during a meeting are based on limited data and tends to describe individual transcription passages based on listening to the speech. We are working on a better account that is based on the multimodal corpus data.

24
Analyse prosody and voice quality of dominance in scenario meetings. We will use the framework developed in the first year to analyse dominance in the AMI Meeting Corpus and we will compare this results with the results already found regarding roles and dialogue acts.

More conversational text-to-speech synthesis. We are collaborating on how to use the rich set of annotations present in the AMI Corpus to develop better conversational speech synthesis and speech signal modification algorithms, by using the appropriate prosody and voice quality features for a given context. We intend to take dominance, role, and dialogue acts into consideration for this work.

In addition, we are currently discussing two possible new network collaborations to be started early in the second year, both of which move the theme into more complex processing. The first will bring more sophisticated machine learning techniques to this kind of data. In it, we will consider the range of tasks for which better learning might be useful and identify individual problems that would most benefit from new techniques, which we would then try. The second is an attempt to port some of the video processing techniques that WP8 has been using on political debates to the AMI Corpus.