

The SSPNet Mobile Corpus: Social Signal Processing Over Mobile Phones

Anna Polychroniou, Hugues Salamin, Alessandro Vinciarelli

University of Glasgow - School of Computing Science
G128QQ Glasgow (UK)
{annap, hasalamin, vincia}@dcs.gla.ac.uk

Abstract

This article presents the SSPNet-Mobile Corpus, a collection of 60 mobile phone calls between unacquainted individuals (120 subjects). The corpus is designed to support research on non-verbal behavior and it has been manually annotated into conversational topics and behavioral events (laughter, fillers, back-channel, etc.). Furthermore, the corpus includes, for each subject, psychometric questionnaires measuring personality, conflict attitude and interpersonal attraction. Besides presenting the main characteristics of the corpus (scenario, subjects, experimental protocol, sensing approach, psychometric measurements), the paper reviews the main results obtained so far using the data.

Keywords: Nonverbal Behavior, Mobile Phones, Social Signal Processing.

1. Introduction

Mobile phones are one of the main channels through which we interact with others. In less than twenty years after their appearance, mobile phones have extended the possibility of contacting (or being contacted by) others to virtually every place and every moment of our day. However, the impact of mobile phones on human-human communication has been studied at the level of large-scale social networks, but largely neglected, to the best of our knowledge, at the level of one-to-one conversations. The SSPNet Mobile Corpus (SMC) aims at filling such a gap and contains 60 dyadic conversations between unacquainted individuals (120 subjects in total).

For each call, the recordings include four audio channels (speaker and microphone for each of the two phones), as well as accelerometer, gyroscope and magnetometer signals for the two phones. In this way, it is possible to study not only what people say and how they say it, but also whether people move and how much. The main reason for adopting the sensors above is that they are all available on most standard mobile phones. Therefore, at least from the signals point of view, the recordings reproduce the data that can be collected with mobile phones in real-world settings. Besides the physical signals, the corpus includes three psychometric questionnaires filled by each subject, namely the Big-Five Inventory 10, the Rahim Organizational Conflict Inventory II and the Interpersonal Attraction questionnaire. The first measures the personality traits, the second measures the conflict handling style and the third measures social and task attractiveness. In this way, it is possible to investigate the relationship, if any, between behaviors detectable in the physical signals and psychometric measurements.

The experimental protocol adopted for the data collection has been designed to reproduce common phone usage scenarios. In particular, the phones were modified only to a minor extent and the recording apparatus was unobtrusive. Furthermore, the subjects were left alone in standard university offices during the experiment. In other words, the conditions were made as close as possible to those encoun-

tered in everyday life. From a content point of view, the subjects were asked to address the Winter Survival Task (WST), a scenario commonly adopted in behavioral experiments.

So far, the corpus was used to investigate the prediction of personality traits and conflict handling style based on nonverbal cues, the automatic detection of vocalizations (fillers and laughter) and the lexical accommodation between speakers. Furthermore, part of the corpus was used as a benchmark in the Computational Paralinguistics Challenge. In this respect, the SMC promises to be a valuable resource for studies in human-human communication and, in particular, Social Signal Processing, the automatic analysis of nonverbal behavior during social interactions.

The rest of the paper surveys previous corpora relevant to the SMC (Section 2.), describes the SSPNet Mobile Corpus (Section 3.), presents the annotation approach (Section 4.), shows the results obtained so far using the data (Section 5.) and draws some conclusions (Section 6.).

2. Previous Work

There is a large number of corpora allowing the investigation of social phenomena including, e.g., the AMI Meeting Corpus (Carletta, 2007), the Mission Survival Corpus (Pianesi and Cappelletti, 2007), the ISL Meeting Corpus (Burger and Sloane, 2004), the NIST Corpus (Garofolo and Przybocki, 2002), the VACE corpus (Chen and others, 2006), the CHIL Corpus (Mostefa et al., 2007), etc. In most cases, these corpora adopt scenarios designed by psychologists and are collected with multiple devices (cameras, microphones, physiological sensors, etc.). In general, the participants are asked to “act” a role they do not play in their real life and the settings are not naturalistic. Furthermore, the corpora often include subjects of different cultures and variable English proficiency.

Since it is not possible to describe in depth all corpora collected in the literature, the rest of this section will focus on those that are most similar to the SSPNet Mobile Corpus, namely the Mission Survival Corpus-2 and the Emergent LEADER Analysis (ELEA) corpus. The similarity between

these corpora might allow one to perform comparative studies.

2.1. Mission Survival Corpus-2

The main similarity between the MSC-2 Corpus and the SSPNet Mobile Corpus is that they are both based on the Winter Survival Exercise (see Section 3. for more details). Furthermore, both corpora include the personality self-assessments of the subjects involved. The main difference is that the subjects of the SSPNet Mobile Corpus do not see one another and cannot rely on the visual feedback they get from their interlocutors. Furthermore, the subjects of the corpus presented in this work are asked to discuss the items of the Winter Survival Exercise sequentially and one after the other. This makes it easier to analyze decision-making processes (the decisions are made at the transition point between one item and the following one) and to distinguish between agreement and disagreement (the participants propose a solution for the task before the call and it is possible to know when they disagree about a given item). The MSC-2 (Mana and Zancanaro, 2007) includes 13 meetings of 4 persons (35 years old, on average, and female in roughly half of the cases). The total length of the meetings is approximately 6.5 hours and the recordings were collected in the CHIL room (Mostefa et al., 2007), a facility equipped with close talk microphones, 6 table and 7 T-shaped array microphones, 5 cameras and 4 webcams. The meetings take place around a circular table without restrictions about position and movements of the participants. The data is annotated in terms of functional and relational roles corresponding to observable behavioral patterns:

- Task Area: Orienter, Giver, Seeker, Recorder and Follower
- Socio-Emotional: Attacker, Gate-keeper, Protagonist, Supporter and Neutral

The behavior was manually annotated in terms of speech activity, and automatically annotated in terms of body movements such as head position, head orientation and fidgeting. Furthermore, the subjects filled two personality questionnaires: Craig's Locus of Control Behaviour scale, the Italian version (Farma and Cortinovis, 2000) and Big Five Marker Scales (Perugini and Di Blas, 2002). This refers to the Extraversion dimension, information about group cohesion in terms of how much a participant felt himself and the others as group members. The subjects provided documentation about their personal solution for the survival exercise scenario.

2.2. Emergent LEADER Analysis

The main contact point between the ELEA and SSPNet Mobile Corpora is that they are both based on the Winter Survival Exercise (see Section 3.). Furthermore, both corpora include psychometric questionnaires measuring personality traits and interpersonal perception. The differences are the same as those observed in the case of the MSC-2 Corpus. Furthermore, while the ELEA Corpus focuses on the dimensions of leadership and dominance, the SSPNet Mobile Corpus focuses on the attitude towards conflict and disagreement (see Section 3. for more information).

The ELEA Corpus (Sanchez-Cortes and Gatica-Perez, 2012) aims at detecting emergent leadership in a meeting using non-verbal behavioural cues. It includes 40 meetings (approximately 10 hours in total) and investigates casual social interactions. It involves 148 participants of average age around 25 years, males in almost two thirds of the cases, who responded to an English/French advertisement posted in a research centre and a business management school in Switzerland. Twenty eight meetings involved 4 people and the remaining twelve only 3. The recordings include audio and video collected with two different set-ups: one static and one portable. In total, 27 meetings were recorded with a portable audio-visual setup, 10 with a static setup and 3 meetings only with microphones.

The corpus includes four questionnaires for each participant:

1. Big Five Personality Trait test, a self-reported version of 60 questions.
2. Personality Research Form, which describes behaviour under power dominance and leadership, 16 questions.
3. Perceived Interaction Score, which measures perception from every other person in the group in terms of perceived leadership, perceived dominance, perceived competence perceived liking, 16 questions. Plus, participants ranked the other parties of the group by dominance.
4. Additional info about age, experience in outdoor activities, winter sports and optional comments on their feelings about the process.

The annotation of the audio derived from the microphone array which can, automatically, annotate the status of every speaker in terms of speaking/non-speaking. Moreover, features were computed, following the binary annotation as: total speaking time per participant, number of turns per participant (where turn is a series of active speaking status), average turn duration per participant and number of interruptions when a participant starts talking in the course of another participant turn.

The visual data provided the information on "*who is looking at whom or what*" (Sanchez-Cortes and Gatica-Perez, 2012) which was utilized to capture attention features such as attention received and given, the ratio of received attention to given attention and number of frames where participant receive attention from all the people in the group simultaneously. Finally, audio-visual features derived from the combination of the unimodal features presented above such as looking while speaking or listening, being looked at while speaking, centre of attention while speaking, etc.

3. The SSPNet Mobile Corpus

The corpus includes 60 calls between dyads of unacquainted individuals (120 subjects in total). The call lengths range between 4 and 35 minutes (11 minutes and 48 seconds, on average) for a total duration of approximately 710 minutes. The length distribution is shown in Figure 1. The

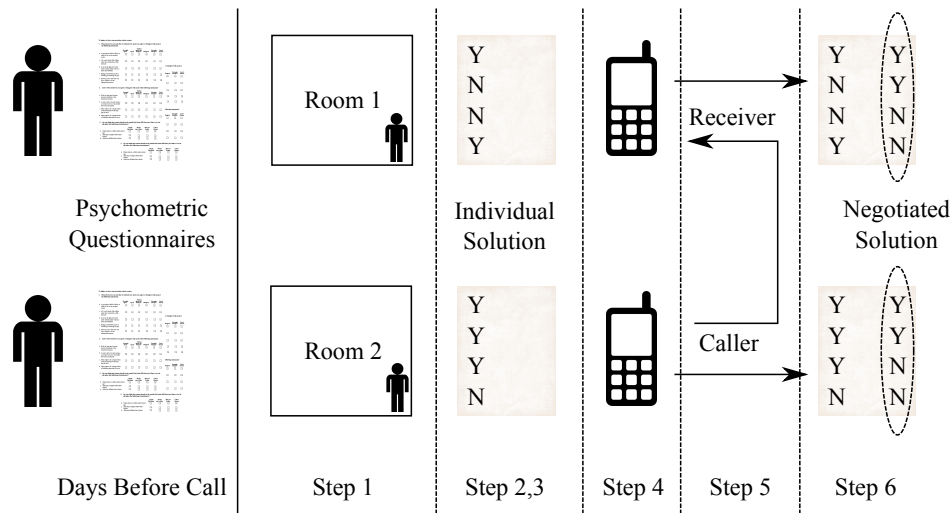


Figure 2: The picture shows the experimental protocol. The subjects fill self-assessment questionnaires in the days before the call. The day of the call, they sit in one of the two offices used for the experiment (step 1), they read the protocol (step 2), they address the WST task individually (step 3), they receive a mobile phone (step 4), they negotiate a common solution during a call (step 5), they deliver a negotiated solution (step 6).

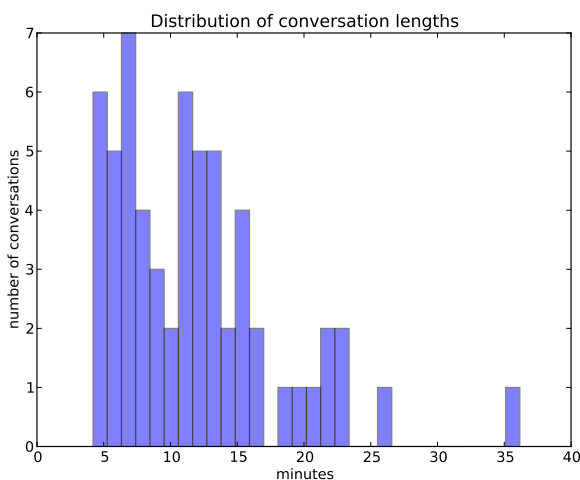


Figure 1: The distribution of conversation length across the database of sixty conversations.

rest of this section provides the most important characteristics of the data, namely information about the participants, scenario, experimental protocol, psychometric questionnaires and sensing apparatus.

3.1. The Subjects

This corpus involves 120 English native speakers between 18 and 64 years old. The median of the participants age is 23 years. The cultural background of the participants is uniform (118 subjects out of 120 hold a British passport). Most of the participants have a university education (the most represented subjects are Psychology and Computing Science) and were recruited at the University of Glasgow: 78 are undergraduate students, 26 postgraduate students or academic staff and 16 are externals, but were students or staff in the past. In terms of gender, 63 are females and 57 males, resulting into 29 male-female dyads, 14 male-male

dyads and 17 female-female dyads.

3.2. Scenario

The participants interact in dyads following a decision-making scenario inspired by the WSE, the Winter Survival Exercise (Volkema and Ronald, 1998). Appendix A shows the document that the subjects receive before being involved in a call. The scenario, the Winter Survival Task (WST) hereafter, is a variant of the original WSE and it requires the participants to discuss over the phone whether 12 items found on the site of a plane crash are useful for survival in a polar area or not. Unlike the WSE, the goal of the participants is not to rank the items, but to identify those that can actually be useful.

The WST ensures that the items are discussed sequentially and each dyad moves from one item to the next only after reaching consensus on whether the item is useful or not. The decisions of the dyad can be compared with a golden rule provided by a survival expert. In order to motivate the participants, there is a reward of £3 every time the dyad identifies as useful an item indicated as such in the golden rule. Symmetrically, there is a £3 penalty when the dyad proposes as useful an item that is not identified as such in the golden rule. In any case, the participants receive a minimum payment of £6.

3.3. Experimental Protocol: the Call

The subjects involved in the data collection, perform all the steps of the following protocol (see Figure 2):

- Step 1: The two subjects are conducted in two separate rooms without encountering one another.
- Step 2: Once in their room, the participants are asked to read the document in Appendix A, where they find the explanation of the task and the list of the 12 items at the core of the WST.

- Step 3: Before starting the call, the subjects fill the left column of the form available at the end of the document in Appendix A with “Yes” or “No” for each item.
- Step 4: The two subjects receive a mobile phone (the same model for both participants).
- Step 5: One of the two subjects, selected randomly, calls the other with the mobile phone provided by the experimenters.
- Step 6: During the call, the subjects discuss item by item and negotiate a common solution (“Yes” or “No”) going through the same document in Appendix A. This is the final outcome of the task.
- Step 7: At the end of the call, the participants fill the Interpersonal Attraction questionnaire (see below).

In the days before the call, the subjects are asked to fill two psychometric questionnaires (see next section). This is a necessary step and the subjects are simply not allowed to perform the experiment without having filled the questionnaires.

3.4. Psychometric Questionnaires

All participants have filled the following psychometric questionnaires:

1. Big Five Inventory-10 (BFI-10) to measure personality traits (Rammstedt and John, 2007)
2. Rahim Organizational Conflict Inventory - II (ROCI-II) to measure conflict handling style (Rahim, 1983)
3. Interpersonal Attraction (IA) questionnaire to measure social and task attractiveness (McCroskey and McCain, 1974).

The first two questionnaires are self-assessments while the third one measures the attractiveness of the interlocutor. Hence, the self-assessments are filled before the call, while the IA is filled after it. Personality is measured because it is one of the main factors behind observable behaviour, conflict handling style because the scenario involves frequent disagreements, and interpersonal attraction because it accounts for the quality of interactions.

3.5. Sensing

Each call corresponds to the following signals:

- Two audio signals (one per phone) capture what the participants say through the phone microphones.
- Two audio signals (one per phone) capture what the participants hear through the phone speakers.
- Two accelerometer signals (one per phone) capture the acceleration along three axes solidal with the phones.
- Two gyroscope signals (one per phone) measure the angular acceleration around the three axes above.

- Two magnetometer signals (one per phone) measure the orientation of the phones with respect to the earth magnetic field.

The sampling rate is 44kHz for the audio signals and 68 Hz for the other ones. Accelerometer, magnetometer and gyroscope are assembled in a small device (a cube of roughly 1.5cm side), the SHAKE (Hughes and O’Modhrain, 2006), attached to the back of the phone. The sensors capture not only what people say, but also the movement of the phone, assumed to be a good approximation of the head movements of the subjects. SHAKE devices and microphones are synchronized with a variable delay of up to 100ms. This is due to the internal working of the phone and delay transmissions through the wireless network used to transfer the data to a storage device.

4. Annotation

By annotation it is meant the segmentation of the audio files into labelled time intervals corresponding to behavioral events (in particular, the most frequent nonverbal cues observed in a conversation) and topics (in particular, the intervals during which each item of the WST is discussed). The data has been annotated with “Transcriber”, a tool publicly available from trans.sourceforge.net. For each call, the annotation was performed over the signal obtained by merging the signal captured with the microphone and the signal emitted by the speaker. In this way, the annotation is performed over data that includes what the subjects both utter and hear. The annotation was performed separately for each of the two phones involved in the call (hence, there are two annotations per call). The main rationale behind this approach is that mobile phones involve transmission delays and, often, an event that takes place at time t for one speaker reaches the other speaker only at time $t + \Delta t$. The double annotation can help to investigate how such delays influence the conversation dynamics.

4.1. Behavioral Events

The annotation (see Figure 3) takes into account the behavioral events most commonly observed during conversations (the two speakers involved in a conversation are indicated as speaker 1 and speaker 2, respectively):

- *Speaking activity*: time intervals during which one speaker is talking.
- *Laughter*: time intervals during which one or both speakers produce vocalizations, like giggling or laughter. An interval is labeled differently depending on whether only speaker 1 is laughing, only speaker 2 is laughing or both speakers laugh together.
- *Overlapping speech*: time intervals during which both speakers are talking at the same time.
- *Back channel*: time intervals during which the listener adopts short vocalizations (e.g., “ah-ah”, “uhm”, etc.) or words (e.g., “yeah”, “yes”, etc.) that manifest attention and/or encouragement but do not aim at grabbing the floor. The label is different depending on whether the back channel is produced by speaker 1 or 2.

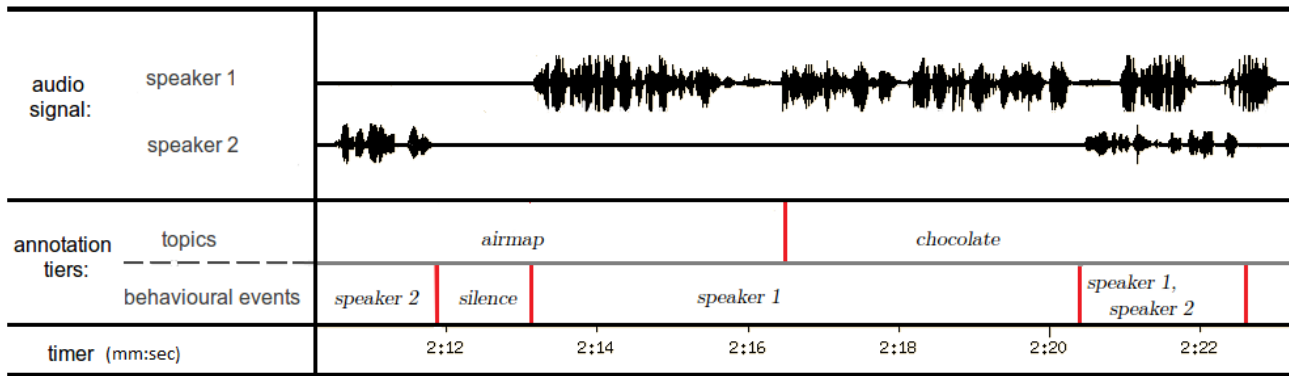


Figure 3: Annotation Scheme

- *Fillers*: time intervals during which speakers utter linguistic vocalizations (e.g. “uhm”, “eh”, etc.) to keep the floor while not saying actual words (often corresponding to hesitations). The label is different depending on whether the filler is produced by speaker 1 or speaker 2.
- *Silence*: time interval when no speech nor vocalization occurs.

4.2. Topics

The conversation content is annotated in terms of the 12 items at the core of the WST (each item corresponds to a topic) and planning activities aimed at solving the task:

- *Items*: time intervals during which one of the 12 items is being discussed, the label is the name of the item (“steel wool”, “axe”, “pistol”, “butter”, “newspaper”, “lighter”, “clothing”, “canvas”, “airmap”, “whisky”, “compass” or “chocolate”).
- *Strategy*: time intervals during which the participants make plans about the best way to address the task and organize the conversation.
- *Other*: time intervals that do not fit in one of the categories above, typically corresponding to small-talk, greetings, mutual introductions, etc.

5. Results

The SSPNet Mobile Corpus was collected to support Social Signal Processing research (Vinciarelli et al., 2009) on phone mediated conversations. In particular, the goal is to make mobile phones more *socially aware* (Pentland, 2005), i.e. more capable of understanding the social phenomena taking place during conversations and, possibly, adapting to them. A few, preliminary results have already been obtained in such a direction.

The behavioral events annotated in the data (see above) have been detected automatically with an approach based on Hidden Markov Models (HMMs) and N -grams (Salamin et al., 2013a). In particular, these latter are shown to improve significantly the performance of the HMMs by taking into account how frequently each of the events follows the others. The experiments were performed

over a fraction of the corpus (segments of 11 seconds centered around laughter and filler events) that was used as a benchmark in the Computational Paralinguistics Challenge (Schuller et al., 2013). In this respect the Corpus has already been shown to be useful to the research community. The prediction of personality traits and conflict handling style, measured with the psychometric questionnaires presented above, was the goal of another approach (Salamin et al., 2013b). The main aspect of the experiments is that both speech and motion data are used in a multimodal approach. Furthermore, the Corpus was used to investigate how lexical accommodation, the tendency of people involved in a conversation to adopt the same words, evolves over time during a conversation (Bonin et al., 2013).

The results above show that the Corpus can help to study not only the phenomena actually targeted in the annotation process (e.g., laughter, fillers and traits), but also those that, while not being annotated, still take place during the conversations (e.g., the lexical accommodation). Current work investigates the influence of the phones on the outcome of the discussions about the items of the WST. The results (still under review at the moment this article is being written) show that subjects receiving the call are more persuasive than those that make it.

6. Conclusions

This paper has presented the SSPNet Mobile Corpus, a collection of 60 mobile phone calls between unacquainted individuals recorded with standard mobile phones. The Corpus has been collected with the main goal of studying human-human communication between people talking via mobile phones, in particular for what it concerns nonverbal communication. However, the Corpus has already been shown to be useful to study linguistic phenomena (in particular lexical accommodation) and interplay between discussion outcomes and role (see Section 5.).

So far, the annotation has focused on the most common nonverbal behavioral cues observed during conversations (in particular, laughter, fillers and back-channel). However, the annotations can be easily enriched with further layers thanks to the use of Transcriber (see Section 3.). In this respect, the Corpus has been designed to be useful beyond the original plans of the authors.

Appendix A

THE SCENARIO

You are member of a rescue team. Your duty is to provide assistance to any person facing dangerous situations in a large area of Northern Canada. You have just received an SOS call from a group of people that survived a plane crash and report on their situation as follows: “Both the pilot and co-pilot were killed in the crash. The temperature is 25°C, and the night-time temperature is expected to be -40°C. There is snow on the ground, and the countryside is wooded with several rivers criss-crossing the area. The nearest town is 32.2 km (~ 20 miles) away. We are all dressed in city clothes appropriate for a business meeting.” The survivors have managed to extract 12 objects from the plane. But they have to leave the site of the accident, carrying only a few objects - the less the better - in order to increase their chances of survival.

THE MISSION

Your mission is to identify the objects most likely to maximize the chances of survival of the plane passengers. The protocol includes two steps:

- Step 1 - Individual Step
You receive a table (p.3) showing the 12 items and you have to decide for each one of them whether it is worth carrying or not. You must write your decision, using YES or NO (YES: they have to carry it, NO: must not carry it), in the column on the left of the table.
- Step 2 - Discussion
You will have a telephone conversation with another member of the rescue team in order to decide together which objects must be carried and what objects must be left in the plane.

As the call is a matter of life and death for the survivors, you will follow an emergency discussion protocol:

1. Consider the first object in the list.
2. Discuss with your colleague whether or not the object must be carried until you make a decision. The decision must be consensual and you can take as much time as you need in order to make the right decision.
3. Write your decision in the column to the right part of the table (p.3): the decision must be the same for both participants.
4. Once you have made a decision, move to the following object and repeat steps 2 and 3.
5. Continue until all objects have been considered and a consensual decision has been made for each one of them.

Please consider the following:

- Discuss one object at a time and move onto the next only after a consensual decision has been made.

- Once a decision has been made, do not go back and change the decision about previous objects. Discuss the objects in the order shown on the attached list.
- Do not interrupt the call until all objects have been discussed and all decisions have been made.

At the end of the conversation you have to return the table with the items, completed with “YES” or “NO” decisions for each item. The results must be the same for both you and your colleague. **The phone call will be recorded.**













A REWARDING SCHEME

You will receive £6 for your participation, but you can significantly increase your reward if you make the right decisions. Some objects are actually necessary and must be carried while others should be left on the crash site:

- You receive £3 extra, each time you decide to carry an item that must actually be carried (a right item).
- You lose £3, each time you decide to carry an item that must not actually be carried.
- You lose £3 for each decision marked on your list that is different from the one of your colleague.

In any case, a payment of £6 is guaranteed for your participation.

TABLE

Your opinion (fill this column BEFORE the call)	Items	Consensus (fill this column DURING the call)
1.	A ball of steel wool	
2.	A small axe	
3.	A loaded 45-caliber pistol	
4.	Can of butter	
5.	Newspapers (one per person)	
6.	Cigarette lighter (without fluid)	
7.	Extra shirt and trousers for each survivor	
8.	6m x 6m (≈20 ft x 20 ft) piece of heavy-duty canvas	
9.	A sectional air map made of plastic	
10.	750 ml of whisky	
11.	A compass	
12.	Family-size chocolate bars (one per person)	

Thanks for your participation!

Acknowledgements

This work was supported in part by the European Commission through the FP7 Network of Excellence SSPNet (GA 241287), in part by the Swiss National Science Foundation via the National Centre for Competence in Research IM2 (Interactive Multimodal Information Management), and in part by the Finnish Ministry for Technological Innovation via the HEI project.

B References

- Bonin, F., De Looze C. Ghosh S. Gilmartin-E. Vogel C. Polychroniou A. Salamin H., Vinciarelli, A., and Campbell, N. (2013). Investigating fine temporal dynamics of prosodic and lexical accommodation. In *Proceedings of Interspeech*.
- Burger, S. and Sloane, Z. (2004). The ISL meeting corpus: Categorical features of communicative group interactions. In *Proceedings of the Meeting Recognition Workshop*.
- Carletta, Jean. (2007). Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. *Language Resources and Evaluation*, 41(2):181–190.
- Chen, L., Rose R. Qiao Y. Kimbara I.-Parrill F. Welji H. Han T. Tu J. Huang Z. Harper M. et al. (2006). VACE multimodal meeting corpus. In *Machine Learning for Multimodal Interaction*, number 3869 in Lecture Notes in Computer Science, pages 40–51.
- Farma, T. and Cortinovis, I. (2000). Un questionario sul “locus of control”: Suo utilizzo nel contesto italiano. *Ricerca in psicoterapia*.
- Garofolo, J., Fiscus J. Martin A. Pallett D. and Przybocki, M. (2002). NIST rich transcription 2002 evaluation: A preview. In *Proceedings of Language Resources and Evaluation Conference*.
- Hughes, S. and O’Modhrain, S. (2006). Shake–sensor hardware accessory for kinesthetic expression. *Proceedings of Enactive*.
- Mana, N., Lepri B. Chippendale P. Cappelletti A. Pianesi F. Svaizer P. and Zancanaro, M.assimo. (2007). Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. In *Proceedings of the Workshop on Tagging, Mining and Retrieval of Human Related Activity Information*, pages 9–14.
- McCroskey, J. C. and McCain, T. A. (1974). The measurement of interpersonal attraction. *Speech Monographs*, 41(3):261–266.
- McOwan, I.). Dev-Audio: Innovative Microphones for Groups.
- Mostefa, D., Moreau N. Choukri K. Potamianos G. Chu S. M. Tyagi A. Casas J. R. Turmo J., Cristoforetti, L., Tobia F., et al. (2007). The CHIL audiovisual corpus for lecture and meeting analysis inside smart rooms. *Language Resources and Evaluation*, 41(3):389–407.
- Pentland, A. (2005). Socially aware, computation and communication. *IEEE Computer*, 38(3):33–40.
- Perugini, M. and Di Blas, L. (2002). Analyzing personality-related adjectives from an eticemic perspective: The Big Five marker scales (BFMS) and the Italian AB5C taxonomy. In De Raad, B. and Perugini, M., editors, *Big Five Assessment*, pages 281–304. Hogrefe und Huber Publishers. Göttingen.
- Pianesi, F., Zancanaro M. Lepri B. and Cappelletti, A. (2007). A multimodal annotated corpus of consensus decision making meetings. *Language Resources and Evaluation*, 41(3):409–429.
- Rahim, A. M. (1983). A measure of styles of handling interpersonal conflict. *Academy of Management Journal*, 26(2):368–376.
- Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1):203–212.
- Salamin, H., Polychroniou, A., and Vinciarelli, A. (2013a). Automatic detection of laughter and fillers in spontaneous mobile phone conversations. In *IEEE International Conference on Systems, Man, And Cybernetics*, pages 4282–4287.
- Salamin, H., Polychroniou, A., and Vinciarelli, A. (2013b). Automatic recognition of personality and conflict handling style in mobile phone conversations. *International Workshop on Image and Audio Analysis for Multimedia Interactive Services*.
- Sanchez-Cortes, D., Aran O. Jayagopi D. Mast-Schmidt M. and Gatica-Perez, D. (2012). Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces*, 6:1–15.
- Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E., Salamin, H., and Polychroniou, A. (2013). The Interspeech 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In *Proceedings of Interspeech*.
- Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social Signal Processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12):1743–1759.
- Volkema, G., Roger J. and Ronald, H. (1998). The influence of cognitive-based group composition on decision-making process and outcome. *Journal of Management Studies*, 35(1):105–121.