Map Task Deviation Scores: A Reconstruction

Justine Reverdy
Computational Linguistics Group
ADAPT Centre
School of Computer Science
and Statistics
Trinity College Dublin,
The University of Dublin
Email: reverdyj@tcd.ie

Hayakawa Akira
Computational Linguistics Group
School of Computer Science
and Statistics
Trinity College Dublin,
The University of Dublin
Email: campbeak@tcd.ie

Carl Vogel
Computational Linguistics Group
Trinity Centre for Computing
and Language Studies
School of Computer Science
and Statistics
Trinity College Dublin,
The University of Dublin
Email: vogel@tcd.ie

***Abstract*—The Human Communication Research Centre Map Task Corpus is a landmark data source for analysis of multi-modal dialogue corpora. The task in this dialogue is the communication by an information giver of a path on a map to an information follower. The corpus includes a measure of task success for each dialogue: the deviation from path scores (the information follower's path from the information giver's path). The original HCRC Map Task works gives information on how the scores for each map were calculated. However, upon analysis and inspection of the corpus materials, we find that it is not possible to reproduce the recorded scores exactly. We report four candidate reconstructions that all appear rational, and note the correlations these lead to with the deviation scores recorded with the corpus. This is important to anyone wishing to reproduce the HCRC method on map-task corpora.**

I. Introduction

Measuring interactional success on its own is not trivial. Cappella (1991) [1] distinguish four approaches: coding, rating, participant judgement, and observer judgement. Coding refers to the assignation of a value to an interactional segment, rating is similar but the values are assigned on a scale for each segments; both coding and rating can be carried out automatically or by trained humans. Judgment approaches do not necessarily require training but have to be done by humans, acting as judges, either judging their own behaviours or someone else behaviours. He notes that each provides different frames of reference, but that instead of searching for a privileged frame of reference, research should be focused at "transforming the results from one frame of reference to another by developing mappings from the more objective measurement frames to those represented by participants and observer judgements" [1, p. 111]. From this statement, one could consider that the verification of the existence of matching patterns from one communication setting to another is desirable, while taking into consideration changes in settings.

How much was understood by each participant of a casual conversation remains dependent on subjective interpretation. However when a specific task that requires coordination is given, the accuracy of completion of the task gives a starting point to assess the quality of the interaction at hand in a more objective manner.

The observations that we present here are about methods of objectively assessing mutual understanding as emerges in task-based dialogue. Objective independent measures of mutual understanding achieved in dialogue may be used in order to assess linguistic signals of mutual understanding and misunderstanding [2]–[9]. Here we focus on a particular multi-modal dialogue corpus, the Human Communication Research Centre (HCRC) Map Task corpus [10]. Our purpose is to analyse the objective measurements of mutual understanding provided with that corpus. We identify rational reconstructions of the terms mentioned in the description of the method used to make those measurements in the hope of being able to replicate the measurements. However, we note that none of these reconstructions produced exactly the same results, although each has a very strong correlation We identify which reconstruction correlates most strongly. We hope that the terms of our reconstruction are such that they may be fully replicated by others on the same corpus, or in use in other corpora based on the same underlying task.

We think that this work is relevant to interaction analysis within cognitive infocommunications, where others also attend to the fine details of corpus construction [11]–[13], as well as the works noted above in estimating mutual understanding on the basis of dialogue transcripts. Dialogue systems may also benefit [14]. As much of cognitive infocommunications dwells on augmenting human cognitive capabilities [15]–[17], it is relevant to provide humans with better means of concluding whether they have understood each other than they evidently have. In the terms of recent conceptualization of interaction analysis [18], [19], this work contributes to refining to methods available to interaction analysis, placing them on sound, replicable footing. The general problem is also relevant to work on navigation and trajectory correction, e.g. in robotics [20].

II. The HCRC Map Task Corpus

The Human Communication Research Centre (HCRC) Map Task corpus consists of 128 dialogues released in 1992 [10]. This corpus uses the map task technique (described below) to elicit spontaneous communicative behaviours in the frame of Human-to-Human task-based interactions. Two subjects per dialogue, with either the role of Information Givers or Information Followers, were each given A3 maps containing landmarks. Almost all participants were native Scottish speakers of English. The Information Giver had a route drawn on the map with a START and a FINISH, and was tasked with guiding the Information Followers through a map containing only landmarks. To add to the difficulty of the task, landmarks from the two maps and their placement differed a little.

The recordings have been divided into eight "quads" of eight participants, themselves divided into two groups of four, one group having eye contact while the other did not. Those divisions allowed a control of three variables susceptible to impact speech variation: eye-contact, gender and familiarity between participants. In the division concerned with the visual cues accessible, half the subjects were able to see their interlocutor's face (i.e., eye contact), while the other half had opaque screens placed between them (i.e., no-eye contact). The subjects could not see their interlocutor's map at any point.

The Information Followers used on average 393.31 tokens per dialogue and the Information Givers 858.10. The participants were 64 in total, with 32 females and 32 males [Reported Gender], that would participate in the task four times, twice as Information Givers and twice as Information Followers, and in each role once with a familiar partner and once with an unfamiliar one.

The task consists in verbal guidance and the participants were told not to use gestures. The Information Givers has to guide the Information Followers along a predefined route, and any *deviations* from that route were assumed to be the result of less successful communication between the two participants, as the subjects were precisely told not to stray from the route.

The precomputed HCRC Map Task corpus deviation score, ranges from 4 (best) to 227 (worst). The higher the score, the more the route deviates from the original route, which is taken as an indication of less successful communication. The relative objectivity behind this reasoning makes the deviation score a particularly good fit to measure success in the frame of this research's methodological development, mostly because it shapes communicative success into a numerical scale.

III. Measuring Communicative Success in a Task

To establish if interlocutors may or may not have understood each other in a dialogue, a means to evaluate their communication have to be agreed upon. All types of face-to-face interactions have their own type of appropriate communicative behaviour, an informal meeting with a friend and a political debate, for example, will have vastly different underlying motives and expectations that will change the communicative style of the participants, influence turn-taking and topic-change. An informal discussion may even have no other motivations than creating social link and the information exchanged have no conscious or exact purposes. In task-based interactions on the other hand, one of the purposes of the exchange, that is to accomplish the task, is explicit, and a measurement of how well a task have been accomplished is available, which gives a means to evaluate the communication. When carrying this evaluation, the type of the task and the medium, will influence communicative style in various ways. The outcome of a task that involves persuasion might be more influenced by the presence of the medium of video, which conveys facial expressions, while problem solving tasks might not be influenced by the presence or absence of a video channel, and if anything its presence could even be a disruption (Whittaker, 2002).

The present work is concerned with one types of task success assessment methods that is used in map tasks to evaluate communicative success: The deviation scores of the map tasks. Those scores are here reconstructed, evaluated, and results from its replication suggest a good reliability of the measure against other possible measure of task success, such as completion time, for the particular task at hand. The deviation scores are a measure that tells by how much a route drawn on a map by an Information Follower deviates from the original route described verbally by an Information Giver.

A. Deviation Scores in the Map Task Technique

Since its first description [21], the map task technique has been used in a wide range of experiments testing for different conditions for which many corpora have been created. Among them the HCRC Map Task [10] is noticeable by its design and accessibility to researchers. The technique was used to create corpora in a variety of English dialects (Scottish, American, Australian), other languages (French Occitan, Italian, Japanese, Portuguese, Dutch, Swedish), in medical context [22], in computer-mediated interactions [13], with the use of avatars [23], and so on. Many studies are using the HCRC map task, and the particular success measure provided by the authors as a variable indicating successful task management. The design of the map task is made so that success in the task is subordinated to successful verbal communication, at least to a certain degree. Half the corpus is in no-eye contact condition with a screen blocking the entire view of the other participant. In the other half – eye contact conditions – a low-height screen was placed to block the view of participants' maps, that could only see each other's face, and they were instructed not to use gestures to communicate (the low-height screens helping with this instruction enforcement). In both conditions, the privileged medium is verbal, and the extent of non-verbal communication is limited to facial expressions but seldom body language. The controlled design of the HCRC Map Task, makes it an exceptional object of analysis for the study of human verbal communicative behaviour, even 30 years after its release. Even by taking into account the evolution of language during that time, it is reasonable to assume that the underlying communicative processes involved in speech dynamics remain fairly similar, if not exactly the same, as they are now.

Following those considerations, much scientific research from fields such as cognitive science, computational linguistics, natural language processing and so on, legitimately used this material in various studies [24]–[28], and also made a use of the measure of success available: the deviation scores [2], [29], [30]. An interesting aspect of this task success measure is that its evaluation gives theoretically the least space to subjective interpretation from the evaluator.

B. Original Pre-Computed Deviation Scores

This work uses the material given by the authors of the HCRC Map Task corpus (The Information Givers' and Followers' maps) to recreate the precomputed score given. Despite its apparent straight-forwardness, they raised a number of issues, even when interested in overall tendencies, that appeared to make the scores worth being replicated to ensure their trustworthiness, in particular to all researchers interested in using the deviation scores as their primary measurement of successful communication. Far from disparaging the measure, we are here rather raising some issues that require resolution and that might be of interest for future researchers in their usage of this method of task success measurement. Replicability of experiments is important for sound scientific research. To build solid foundations for hypothesis testing that will in turn lead to the construction of the theories that constitute scientific knowledge, the capacity to replicate experimental results is crucial in all domains. In particular in domains closely related to psychology such as computational linguistics and cognitive sciences, where the difficulty to replicate results have been more than often pointed out [31].

C. Reconstruction of the Deviation Scores

The reason for the redesign of path deviation counting methods was mainly that the same results were not found after few attempts at replicating the scores of different maps; even when carefully reverse engineering the A3 to A4 specifications and the 1 cm grid overlay, that should have led to the same scores. Similar, but not exact scores were found, which led to the question of replicabilty and most importantly reliability of the scores meant to be used as a starting point to evaluate communication in human behaviour in subsequent experiments.

The scores needed to be accurate and possible to replicate with precision. A number of issues appeared, such as the poor quality of the scanned Information Follower maps and their apparent distortion during the scanning process of the completed maps. It was indeed difficult to make match the given Information Followers original maps that were the templates for the maps and the actual Information Followers maps that had the results routes drawn on them. The software Adobe Illustrator was used to respect precisely the instructions given (A3 grids over A3 Information Followers/Information Givers maps). Following the instructions resulted in the definition of each possible cases in which a square can be in 4 different methods of counting, from the most restricted possibility to the broadest interpretation.

A number of descriptions of the method created to count deviation scores can be found in the literature and were used to reconstruct those scores from the original maps given with the HCRC Map Task corpus. In a widely cited work as originating the deviation scores [10], no mention of a scoring method is found. This work uses the map task technique to create 170 dialogues of children between 5 and 13 (in three groups), and mention that this technique was created by Brown, Anderson, Yule & Shillcock [21]. No mention of deviation scores is made in that publication either. The results in that publication are concerned with the proportions of definite and indefinite article usage among the different categories of speakers — as it is the declared aim of the research article — but not in relation with the participants success in the task.[1] The book cited by Anderson et al., '*Teaching Talk: Strategies for Production and Assessment.*' [21], however, does mention an assessment method, even if it is not a deviation score per se. The map task was originally designed to assess pupils' information transfer skills, and was tested among other tasks such as wiring-broad task or story summarizing. As the corpus creators highlight, despite the requirement that the route is to be drawn as precisely as possible in order to avoid potential danger, "it is not draughtsmanship which is being assessed, but the ability to recognise that the other person needs to be told whether to go right or left, or up or down at crucial parts of the map" [21, p. 111]. However, the relative precision of the reproduction of the route remain key to the reliability of the measure. The deviation scores are described as the centimetre square difference between the route on the map of the Information Giver and the Information Follower, with the map divided into a grid of one centimetre squares.

The description given on the HCRC Map Task website[2] was used to define four methods of counting, depending on the area between the route drawn by the Information Follower and the original route, each described in Table I. Each grid of 1 cm squares has been placed to overlay the Information Followers maps from the top left. The grids and the Information Follower maps are subsequently placed on top of the Information Giver maps by trying to match the starting cross first, then the different landmarks. In some cases, this operation required slight distortions for a better match between the maps, as it seemed the Information Followers maps provided suffered distortions in the scanning process. How to estimate the distance between the two lines and decide if a square of the grid that overlay the two lines should be included in the score, or in other terms: 'is the area between the two routes large enough to be counted as a deviation?'. In an attempt of precision and after an observation of the real resulting maps, eighteen possible situations have been identified (involving forty-five squares) in which the lines and overlays could be.[3] This is illustrated in Figure 1.

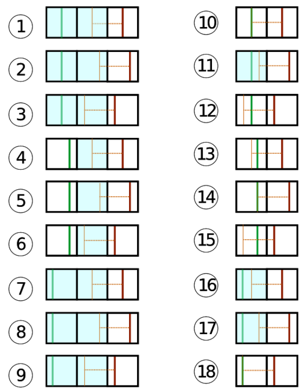
TABLE I: Description of Deviation Scores Counting Methods.

	Description
Method 1	Counting squares of which the area is more than 50% between the 2 routes.
Method 2	In addition to Method 1, include the squares that cover the Information Followers path but of which the area is not more than 50% between the two routes.
Method 3	In addition to Method 2, count the squares including the Information Givers route when more than 50% of the area is between the two routes.
Method 4	In addition to Method 1, count the squares including the Information Givers route when more than 50% of the area is between the two routes.

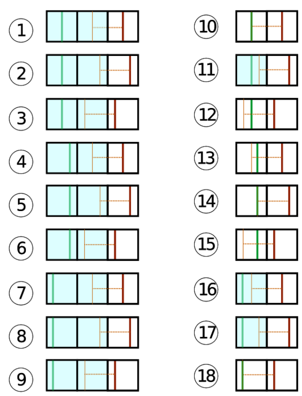
[1]No mention in the study of a task score or any variable linked to task success.

[2]http://groups.inf.ed.ac.uk/maptask/maptask-description.html (Last consulted: 01/07/2020)

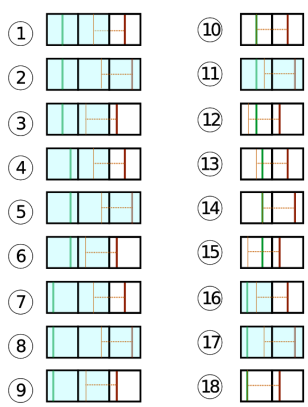
[3]This number of eighteen situations comes from the three possible positions in which a line can be in terms of division of the space inside a square; either the area of the square is divided into 50% of the space on each side of the line, or the percentage is higher/lower on each side. (As there are two lines that can be in three positions in each square, with possibly an entire square in between this, results in: 2 lines times three positions times three situations equals eighteen.)



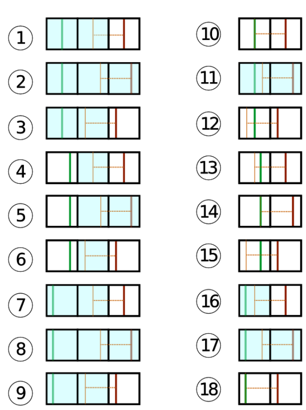
(a) Method 1: 18 Squares



(b) Method 2: 21 Squares



(c) Method 3: 26 Squares

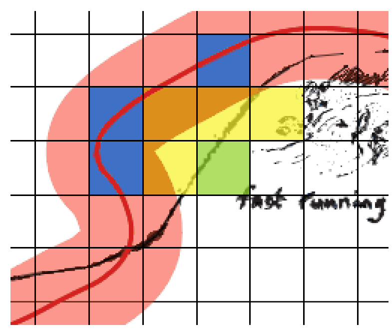


(d) Method 4: 23 Squares

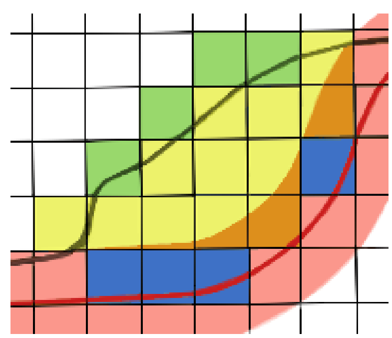
Fig. 1: Possible cases for each square in the Methods of counting (18 possibilities)

In the eighteen possible configurations in which the drawn route can be positioned with respect to the original route that have been isolated, each method kept a certain number of squares. Respectively 18 squares for Method 1, 21 for Method 2, 26 for Method 3 and 23 for Method 4. in which the red line represents the original route (right), the yellow line the distance of one centimetre (centre), and the green line the route drawn by the Information Followers (left). This can be interpreted as the distance from which the path starts to be considered deviant, Method 1 being the least inclusive (the strictest interpretation of the described method) and Method 3 the most inclusive (the broadest interpretation).

The methods of counting have been applied to the 128 maps of the HCRC Map Task. The results of these counts can be found in Table III. An example of counting on the real maps, to exemplify the possibilities of different interpretations if no clear definition of the counting method is given, can been seen in Figure 2. The red line represents the original route, and the black line the route drawn by the Information Followers, the light red area (which is light brown when seen under yellow) represents one centimetre distance from the original route. The squares in yellow are counted using method 1, the squares in green are the addition created by following counting method 2, the squares in blue are added for method number 3, finally method number 4 retains the squares in blue and yellow only.



(a) Example 1



(b) Example 2

Fig. 2: Counting Methods real example from HCRC maps

To assess which method has the minimum distance with the HCRC Deviation Score, a series of Correlation Tests have been applied on the methods (see Table II). Method 4 is the closest to the HCRC pre-computed method, as it is the method showing the highest correlation with the original pre-computed scores.

TABLE II: Pearson Correlation Coefficients for Pre-Computed Deviation score and given Counting Methods

Scores compared	Correlation coef. R
HCRCDevS,M1	0.944509
HCRCDevS,M2	0.9426531
HCRCDevS,M3	0.9492633
HCRCDevS,M4	0.9501321
M1,M2	0.9938681
M1,M3	0.9922469
M1,M4	0.9956481
M2,M3	0.9966103
M2,M4	0.9886506
M3,M4	0.9946619

The results show a very high correlation between the pre-computed deviation scores and the recounted ones, which still indicates the relative precision of the original scores, even if it was needed to verify their accuracy after noticing the differences.

D. Other Map Task Scoring Systems and Limitations

Time of completion is also used as measure of success for map tasks [30], however no correlation between deviation scores and time were found in the HCRC Map Task (Pearson correlation coefficient, r = -0.08) and note that participants in the map task were told to be accurate (Teaching Talk, 1984) and not to finish as fast as possible, which makes time completion a less appropriate measure for this particular task. In her studies examining the structuring principles of task-oriented dialogues, Bethan L. Davies uses the HCRC Map Task for which she has created an alternative scoring system [26], [32]: the Incorrect Entity score. She identifies two disadvantages of the deviation scores method. Her first point is that (p.102) "[...] estimating portions of grid squares is not straightforward: it is inevitable that the section being calculated will not always contain whole grid squares." This first point is addressed with the above described methods of counting, which define all possible situations in which a square can be – and the problem of subjectivity that arises when including a square in the counting or not is blur – is greatly reduced. Her second point remains however valid despite the counting method proposed in this section: as it is not draughtsmanship that is measured, Information Followers that negotiate correctly the landmarks but do not accurately follow the original route are penalised by the method. They were told to be accurate, as explorer that had to follow the route described precisely as it is the only "safe" route, but not to the centimetre precision that the deviation scores might suggests. Despite this shortcoming, the deviation scores remain pertinent measures in the frame of this study by its objectiveness and accuracy, above time completion.

IV. CONCLUSION

This paper has drawn attention to the fact that it is difficult to replicate the computation of the path deviation scores released with the HCRC Map Task corpus. Four candidate reconstructions of that method have been presented and tested in correlation with the deviation scores released with that corpus. None is a perfect match. However, hopefully each is fully replicable in future work with the HCRC corpus and with follow-on corpora that have followed the HCRC model.

TABLE III: Deviation Score per Method of Counting

DialID	M1	M2	M3	M4	HCRC	DialID	M1	M2	M3	M4	HCRC
q1nc1	84	99	119	104	78	q1ec1	148	170	197	175	135
q1nc2	221	237	269	253	204	q1ec2	287	321	363	329	227
q1nc3	55	63	69	61	40	q1ec3	92	117	133	108	74
q1nc4	56	66	81	71	53	q1ec4	129	147	168	150	152
q1nc5	51	62	72	61	35	q1ec5	121	151	172	142	105
q1nc6	35	46	56	45	34	q1ec6	52	59	67	60	53
q1nc7	28	32	39	35	18	q1ec7	121	145	164	140	104
q1nc8	95	105	115	105	69	q1ec8	97	130	143	110	85
q2nc1	51	63	69	57	51	q2ec1	128	141	162	149	120
q2nc2	100	116	131	115	90	q2ec2	104	125	138	117	97
q2nc3	56	63	79	72	44	q2ec3	115	129	161	147	105
q2nc4	101	107	124	118	104	q2ec4	119	138	154	135	142
q2nc5	51	61	81	71	41	q2ec5	61	67	83	77	66
q2nc6	94	123	144	115	99	q2ec6	78	92	111	97	89
q2nc7	30	36	42	36	20	q2ec7	40	44	48	44	28
q2nc8	27	37	44	34	19	q2ec8	72	85	93	80	58
q3nc1	232	263	281	250	105	q3ec1	159	193	220	186	140
q3nc2	96	115	130	111	114	q3ec2	197	239	261	219	191
q3nc3	55	68	84	71	56	q3ec3	24	33	40	31	28
q3nc4	42	49	60	53	45	q3ec4	133	145	166	154	154
q3nc5	129	145	176	160	139	q3ec5	29	35	38	32	20
q3nc6	86	100	122	108	99	q3ec6	75	79	102	98	89
q3nc7	66	85	92	73	55	q3ec7	77	93	102	86	57
q3nc8	51	58	64	57	42	q3ec8	50	61	69	58	37
q4nc1	83	106	121	98	74	q4ec1	200	232	249	217	204
q4nc2	112	138	159	133	83	q4ec2	127	144	157	140	117
q4nc3	73	77	97	93	75	q4ec3	35	39	49	45	11
q4nc4	31	33	45	43	45	q4ec4	72	88	111	95	73
q4nc5	30	37	46	39	21	q4ec5	26	29	38	35	26
q4nc6	39	41	53	51	37	q4ec6	37	42	57	52	49
q4nc7	40	49	55	46	28	q4ec7	52	67	78	63	43
q4nc8	39	47	52	44	20	q4ec8	48	58	67	57	30
q5nc1	75	107	121	89	64	q5ec1	139	168	191	162	146
q5nc2	262	293	324	293	201	q5ec2	197	246	267	218	187
q5nc3	59	66	85	78	29	q5ec3	78	94	105	89	66
q5nc4	92	102	122	112	90	q5ec4	6	7	8	7	4
q5nc5	36	44	50	42	17	q5ec5	45	52	56	49	28
q5nc6	78	93	110	95	65	q5ec6	40	52	65	53	38
q5nc7	132	162	186	156	108	q5ec7	157	209	221	169	154
q5nc8	44	54	65	55	28	q5ec8	132	147	165	150	102
q6nc1	83	97	111	97	83	q6ec1	55	69	78	64	38
q6nc2	75	87	94	82	56	q6ec2	112	119	130	123	81
q6nc3	170	209	240	201	161	q6ec3	85	105	131	111	97
q6nc4	121	156	181	146	108	q6ec4	47	51	64	60	35
q6nc5	194	242	271	223	178	q6ec5	63	72	89	80	60
q6nc6	47	49	74	72	52	q6ec6	78	96	108	90	32
q6nc7	68	83	94	79	52	q6ec7	51	58	61	54	54
q6nc8	167	193	209	183	145	q6ec8	83	86	91	88	67
q7nc1	179	228	234	185	157	q7ec1	101	122	133	112	86
q7nc2	63	74	90	79	56	q7ec2	12	13	17	16	7
q7nc3	99	111	130	118	104	q7ec3	33	52	60	41	27
q7nc4	20	20	25	25	35	q7ec4	51	54	60	57	57
q7nc5	41	47	63	57	34	q7ec5	19	27	38	30	24
q7nc6	19	21	26	24	26	q7ec6	9	9	9	9	7
q7nc7	56	72	78	62	48	q7ec7	45	51	60	54	27
q7nc8	43	52	59	50	53	q7ec8	46	53	65	58	38
q8nc1	43	50	68	61	25	q8ec1	51	60	76	67	37
q8nc2	152	166	190	176	112	q8ec2	132	164	194	162	135
q8nc3	52	66	78	64	54	q8ec3	12	13	19	18	11
q8nc4	48	58	76	66	76	q8ec4	43	52	69	60	54
q8nc5	45	63	76	58	40	q8ec5	18	20	25	23	25
q8nc6	53	68	88	73	62	q8ec6	37	44	61	54	47
q8nc7	64	74	85	75	47	q8ec7	51	60	73	64	43
q8nc8	58	65	67	60	49	q8ec8	32	37	42	37	19

Method 4 is the one we recommend for further applications. In conclusion, the method for reconstructing deviations score presented in this document point out the accuracy of deviations scores as measures to evaluate understanding between participants, while bringing more precision on how to create them for anyone interested in human communication evaluation in general, and map tasks in particular.

ACKNOWLEDGMENT

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

REFERENCES

[1] J. N. Cappella, "Mutual Adaptation and Relativity of Measurement," in *Studying Interpersonal Interaction*, B. M. Montgomery and S. Duck, Eds. Guilford Press, 1991, vol. 1, pp. 103–117.

[2] D. Reitter and J. D. Moore, "Predicting Success in Dialogue," in *ACL 2007 — Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague, Czech Republic: Association for Computational Linguistics, 2007, pp. 808–815.

[3] C. Vogel and L. Behan, "Measuring synchrony in dialog transcripts," in *Cognitive Behavioural Systems: Lecture Notes in Computer Science*, Esposito A., Esposito A.M., Vinciarelli A., Hoffmann R., Müller V.C., Ed. Springer, Berlin, Heidelberg, 2012, vol. 7403, pp. 73–88.

[4] C. Vogel, "Attribution of Mutual Understanding," in *Journal of Law and Policy*, vol. 21.2. HeinOnline, 2013, pp. 377–420.

[5] D. Reitter and J. D. Moore, "Alignment and task success in spoken dialogue," *Journal of Memory and Language*, vol. 76, pp. 29–46, 2014.

[6] J. Reverdy and C. Vogel, "Linguistic Repetitions, Task-based Experience and A Proxy Measure of Mutual Understanding," in *Proceedings of CogInfoCom 2017: the 8th IEEE International Conference on Cognitive InfoCommunications*. Debrecen, Hungary: IEEE, September 2017, pp. 395–400.

[7] ——, "Measuring Synchrony in Task-Based Dialogues," in *Proceedings of INTERSPEECH'17: the 18th Annual Conference of the International Speech Communication Association*. Stockholm, Sweden: ISCA, 2017, pp. 1701–1705.

[8] J. Reverdy, A. Hayakawa, and C. Vogel, "Alignment in a multimodal interlingual computer-mediated map task corpus," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, H. Koiso and P. Paggio, Eds. Paris, France: ELRA, Workshop on Language and body in real life & Multimodal Corpora 2018, 2018, pp. 55–59.

[9] J. Reverdy, M. Koutsombogera, and C. Vogel, "Linguistic repetition in three-party conversations," in *Neural Approaches to Dynamics of Signal Exchanges*, A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero, Eds. Singapore: Springer Singapore, 2020, pp. 359–370.

[10] A. H. Anderson, M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert, "The HCRC Map Task Corpus," *Language and Speech*, vol. 34, no. 4, pp. 351–366, 1991.

[11] L. Hunyadi, K. Bertók, T. E. Németh, I. Szekr ényes, Á. Abuczki, G. Nagy, N. Nagy, and P. N. a nd A. Bódog, "The outlines of a theory and technology of human-computer interaction as represented in the model of the hucomtech project," in *2011 2nd International Conference on Cognitive Infocommunications (CogInfoCom)*, 2011, pp. 1–5.

[12] L. Hunyadi, "On multimodality in the perception of emotions from materials of the huco mtech corpus," in *2015 6th IEEE International Conference on Cognitive Infocommunication s (CogInfoCom)*, 2015, pp. 489–492.

[13] A. Hayakawa, S. Luz, L. Cerrato, and N. Campbell, "The ILMT-s2s Corpus — A Multimodal Interlingual Map Task Corpus," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds. Paris, France: European Language Resources Association (ELRA), 2016.

[14] S. Ondáš, J. Juhár, M. Pleva, P. Ferčák, and R. Husovskỳ, "Multimodal Dialogue System with NAO and VoiceXML Dialogue Manager," in *Proceedings of CogInfoCom 2017: the 8th IEEE International Conference on Cognitive InfoCommunications*. Debrecen, Hungary: IEEE, 2017, pp. 439–444.

[15] P. Baranyi and A. Csapo, "Cognitive infocommunications: Coginfocom," in *2010 11th International Symposium on Computational Intelligence a nd Informatics (CINTI)*, 2010, pp. 141–146.

[16] P. Baranyi and Á. Csapó, "Definition and Synergies of Cognitive Infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

[17] P. Baranyi, Á. Csapó, and P. Várlaki, "An overview of research trends in CogInfoCom," *IEEE 18th International Conference on Intelligent Engineering Systems INES 2014*, 2014.

[18] C. Vogel and A. Esposito, "Linguistic and behavior interaction analysis within cognitive infocommunications," in *10th IEEE Conference on Cognitive Infocommunications*, 2019, pp. 47–52.

[19] ——, "Interaction analysis and cognitive infocommunications," *Infocommunications Journal*, vol. 12, no. 1, pp. 2–9, 2020.

[20] M. Pleva and A. Cizmar, "Car trajectory correction and presentation using google maps," *Communications-Scientific letters of the University of Zilina*, vol. 17, no. 1, pp. 121–125, 2015.

[21] G. Brown, A. Anderson, R. Shillcock, and G. Yule, *Teaching talk: Strategies for production and assessment*. Cambridge University Press, 1985.

[22] E. G. Bard, C. Sotillo, A. H. Anderson, H. S. Thompson, and M. M. Taylor, "The DCIEM Map Task Corpus: Spontaneous dialogue under sleep deprivation and drug treatment," *Speech Communication*, vol. 20, no. 1-2, pp. 71–84, 1996.

[23] E. L. Clayes and A. H. Anderson, "Real faces and robot faces: The effects of representation on computer-mediated communication," *International Journal of Human-Computer Studies*, vol. 65, no. 6, pp. 480–496, 2007.

[24] J. Carletta, A. Isard, S. Isard, J. C. Kowtko, G. Doherty-Sneddon, and A. H. Anderson, "The Reliability of a Dialogue Structure Coding Scheme," *Computational Linguistics*, vol. 23, no. 1, pp. 13–32, 1997.

[25] C. F. Sotillo, "Phonological reduction and intelligibility in task-oriented dialogue," Ph.D. dissertation, University of Edinburgh, 1997.

[26] B. L. Davies, "An empirical examination of cooperation, effort and risk in task-oriented dialogue," Ph.D. dissertation, University of Edinburgh, 1997.

[27] J. S. Pardo, "On phonetic convergence during conversational interaction," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006.

[28] K. P. Truong and D. Heylen, "Measuring prosodic alignment in cooperative task-based conversations," in *Proceedings of INTERSPEECH'12: the 13th Annual Conference of the International Speech Communication Association*. Portland, Oregon, USA: ISCA, 2012, pp. 843–846.

[29] M. Colman and P. Healey, "The Distribution of Repair in Dialogue." in *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, L. Carlson, C. Hoelscher, and T. F. Shipley, Eds. Austin, Texas, USA: Cognitive Science Society, 2011, pp. 1563–1568.

[30] C. D. Rothwell, "Recurrence quantification models of human conversational grounding processes: Informing natural language human-computer interaction," Ph.D. dissertation, Wright State University, 2018.

[31] J. B. Asendorpf, M. Conner, F. De Fruyt, J. De Houwer, J. J. A. Denissen, K. Fiedler, S. Fiedler, D. C. Funder, R. Kliegl, B. A. Nosek, M. Perugini, B. W. Roberts, M. Schmitt, M. A. G. van Aken, H. Weber, and J. M. Wicherts, "Recommendations for Increasing Replicability in Psychology," *European Journal of Personality*, vol. 27, no. 2, pp. 108–119, 2013.

[32] B. L. Davies, "Testing dialogue principles in task-oriented dialogues: An exploration of cooperation, collaboration, effort and risk," *Leeds Working Papers in Linguistics and Phonetics*, vol. 11, pp. 30–64, 2006.