



Providing Objective Metrics of Team Communication Skills via Interpersonal coordination Mechanisms

Céline De Looze¹, Brian Vaughan², Finnian Kelly³, Alison Kay⁴

¹Phonetics and Speech Laboratory, Trinity College Dublin, Ireland

²Dublin Institute of Technology, Ireland

³Center for Robust Speech Systems, University of Texas at Dallas, USA

⁴Centre for Innovative Human Systems, School of Psychology, Trinity College Dublin

deloozec@tcd.ie, brian.vaughan@dit.ie, finnian.kelly@utdallas.edu, Alison.Kay@tcd.ie

Abstract

Being able to communicate efficiently has been acknowledged as a vital skill in many different domains. In particular, team communication skills are of key importance in the operation of complex machinery such as aircrafts, maritime vessels and such other, highly-specialized, civilian or military vehicles, as well as the performance of complex tasks in the medical domain. In this paper, we propose to use prosodic accommodation and turn-taking organisation to provide objective metrics of communication skills. To do this, human-factors evaluations, via a coordination Demand Analysis (CDA), were used in conjunction with a dynamic model of prosodic accommodation and turn-taking organisation. Using conversational speech from airline pilots involved in a collaborative task (decision-making exercise), our study reveals that interpersonal coordination mechanisms are indicative of human evaluation of pilots' communication skills. We discuss our results in terms of relevance for training simulation for personnel in safety or mission critical environments.

Index Terms: prosodic accommodation, turn-taking organisation, communication skills, aviation

1. Introduction

1.1. The importance of team communication skills

Being able to communicate efficiently has been acknowledged as a vital skill in many different domains. In particular, team communication skills have been highlighted as a core challenge in the operation of complex machinery such as aircrafts, maritime vessels and such other, highly-specialised, civilian or military vehicles, as well as the performance of complex tasks in the medical domain.

The ability to work together with partners of a team has been pointed out as an essential skill to enhance communication and information exchange and to improve team effectiveness and cohesiveness [1, 2]. In this context, communication skills training programs have become widespread over the past decade, with the military and aviation communities in particular having made large investments in this field for better understanding the benefits derived from improved team skills.

For instance, with reference to the specific field of aircraft operational training, the technical and operational parameters of aircrafts are well understood and defined: aircrafts contain complex data measuring and processing systems for tracking and recording operational parameters, and check that they are maintained within defined parameters in order to identify possible operational or technical errors. Whilst pilots are required to

have the technical knowledge to fly and manage their aircraft, they are also required to have effective communication skills in order to operate safely and efficiently in a mission-critical environment. Pilot error can be costly and terminally dangerous, thus effective communication skills need to be maintained in flight to ensure that no operational errors can occur due to a lack of understanding or communicative deterioration between a pilot and the rest of the crew, particularly a co-pilot.

Accordingly, numerous training protocols exist to train pilots and crew in the non-technical aspects of flying a plane: Multi Crew Cooperation (MCC), Jet Orientation Course (JOC) and Crew Resource Management (CRM) [3, 4, 5]. These training protocols are currently assessed by trainers, with no method of objective measurement being currently possible. Considering that the core competencies of aviation can broadly be described as Aviate, Navigate and Communicate, this is potentially problematic in that the assessment of communication is rendered a non-technical 'soft' skill alongside the objective technical skills of aviation and navigation. It is the authors' contention that recent developments in the area of speech analysis, specifically prosodic accommodation and turn-taking organisation, enable the subjective assessment of communication skills to have an added objective dimension.

A difficulty lies in how to evaluate team communication skills in simulation-based training, and what criteria to evaluate. One proposal is to evaluate both team performance and team effectiveness [6]: the first concept refers to the behavioural cognitive and affective processes which teams engage in, for coordinating their interactions toward shared goals. The second concept corresponds to the results, or success, of the performance. In this dual conceptualisation, it is proposed that evaluations should not only capture the outcome of team performance, but also how the team achieves the outcome. The concept of team performance specifically deals with how individuals within a team dynamically adjust, coordinate and cooperate their behaviour to one another. In this context, we propose that interpersonal coordination mechanisms (namely prosodic accommodation and turn-taking organisation) can provide insights of team performance.

1.2. Interpersonal coordination mechanisms and team performance

Spoken interaction is a joint activity where all participants are involved in the co-construction of meaning and in the establishment and maintenance of social relationships. Turn-taking organisation and accommodation mechanisms are instances of

such coordination or cooperation among individuals, and organises and establishes when to talk, listen and take turns. Accommodation mechanisms are utilised by interlocutors to take similar facial expressions, body postures and gestures [7, 8, 9], and adapt their way of speaking to that of their interlocutor at different linguistic levels: semantic, syntactic, phonological, phonetic and prosodic [10, 11, 12, 13, 14]. In particular, speakers adapt their prosodic characteristics (e.g. intonation contours, pitch range, speech rate, voice intensity) to one another. While there are a number of terms used to describe this phenomenon (convergence, synchrony, mimicry), the most pertinent in the context of this paper is prosodic accommodation.

Interpersonal coordination mechanisms are important aspects of social interaction as they facilitate comprehension, and increase speakers social engagement and affinity. In [14], using telephone conversations in Japanese, we have shown that the higher the level of prosodic accommodation, the higher the level of engagement and the greater the affinity between the interlocutors. Prosodic accommodation is a subtle mechanism, which requires cognitive, linguistic, physiological and psychosocial skills. If these are deficient - for instance under circumstances of high-stress environments or in the case of lack of linguistic knowledge - they may prevent the realisation of accommodation mechanisms. As a result, difficulties in interacting may arise and affect the quality of an interaction, in terms of information exchange and inter-personal relationships.

In this paper, we compare the data obtained from our dynamic model of interpersonal coordination mechanisms with the results of an established Human Factors method for measuring communication skills, adaptability, and teamwork: coordination Demand Analysis (CDA) [15]. The results substantiate the veracity of our model as a method of objectively measuring team communication skills.

2. Experiment

2.1. Set-up

In collaboration with the Irish pilots union IALPA, a group of Irish commercial pilots were assembled, with the goal of obtaining insights into inter-pilot communication. Twelve pilots volunteered to take part in a research session held at a hotel meeting room at Dublin Airport. The pilots took part in a collaborative task (desk-top decision-making exercise) for ten minutes while their vocal interactions were recorded and two human factors researchers assessed their overall Crew Resource Management (CRM) abilities. These recordings were then analysed and compared to the results of the CRM analysis in order to determine whether dynamic model of interpersonal coordination mechanisms captured salient aspects of CRM assessment. Figure 1 details the set-up of the equipment and people on the day.

2.2. Speech data

The speech data for this study consisted of six 10-minute dialogue interactions between Irish commercial pilots (twelve participants, six teams in total). The teams were presented with 15 images of items necessary to survive an imaginary shipwreck scenario. Figure 2 details the scenario and screen as presented to the participants. The pilots were told they had 10 minutes to correctly order the items to achieve a score of 15, getting one point for each correct decision and losing a point for each incorrect decision. The pilots were also told that they would get negative audio feedback (a warning beep) every time they would make an incorrect decision. For this experimental design, pilots

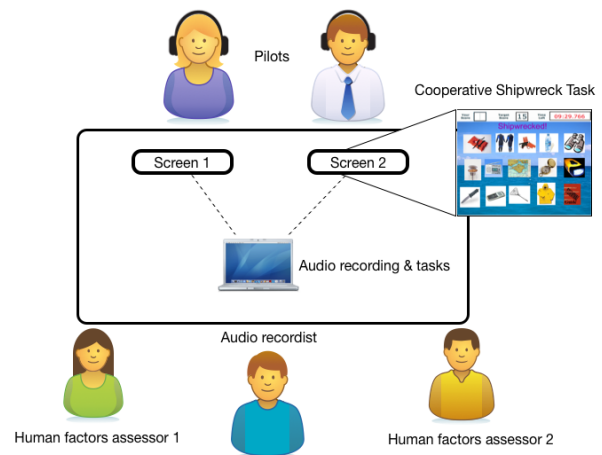


Figure 1: Illustration of the layout of the experimental setup and the personnel involved

actually received positive scores up until the five minute mark after which audio beeps were sounded at regular intervals. Each time an audio beep was sounded the time of the beep was noted for later analysis. The speech audio was recorded using two DPA lapel microphones that ran into an Avid M-Box connected to a laptop running Pro-Tools software. All audio was recorded at a 44.1KHz/16 bit resolution. Each pilot was fitted with a headset to reduce any external noise distractions and to ensure they could clearly hear each other and the audio warnings during the task.

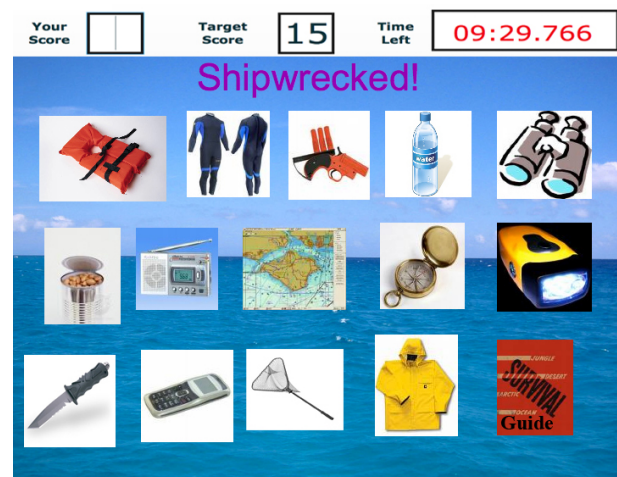


Figure 2: Example of the basic shipwreck task given to the participants

2.3. Annotation and measurement

2.3.1. Speech data annotation

The data was annotated in terms of speech units (or Inter-Silence Units, ISUs) and silences automatically. A binary voice activity detection (VAD) was carried out on both speaker channels for each dyadic interaction, using the VAD algorithm proposed in [16]. The signal was labelled as speech or silence

based on the energy over 250 ms frames. Pauses, gaps, no-gap-no-overlaps (NGNO) and transition overlaps (TOV) were also annotated and determined from the speech/silence annotation. A schematic output of the annotation is shown in Fig. 3. The Matlab signal processing software was used for the annotation and analyses.

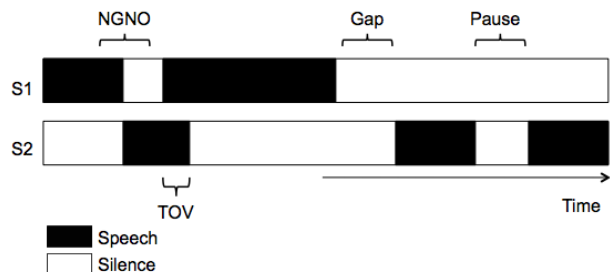


Figure 3: Schematic of a dyadic conversation between speaker 1 (S1) and speaker 2 (S2), illustrating occurrences of pauses, gaps, no-gap-no-overlaps (NGNO) and transition overlaps (TOV).

2.3.2. Speech measurements

Turn-taking organization. Turn-taking organisation is measured as a set of speech timing features based on the annotation described in section 2.3.1: (i) global amount of speaking/ silent times (given in %), (ii) global amount of overlaps (given in %), (iii) number and duration (in seconds) of turn transitions (gaps, overlaps,), (iv) number and duration of silent pauses, (v) number and duration of ISUs, (vi) number and duration of turns and (vii) number and durations of overlaps (which are not transition overlaps). Features are extracted per team and per speaker.

Prosodic accommodation. The method developed by the two first authors [14] was used to measure prosodic accommodation. First, prosodic features (pitch and energy) are extracted per speaker using a time aligned moving average method. F0/Energy median and standard deviation about the mean are extracted for every moving window of 30 s duration. These features, as well as the durations of pauses, gaps, transition overlaps, ISUs, turns and overlaps, are then used as input features to calculate the level of accommodation in pitch, energy and timing respectively. The level of accommodation per team is measured using the Spearman correlation coefficient $\rho \in [-1, 1]$. Large $\rho_{xy} \gg 0$ indicate a high level of accommodation. Small $\rho_{xy} \ll 0$ indicate a low level of accommodation. The level of prosodic accommodation per speaker is computed using a normalised area-under-the-curve (AUC) calculation for the prosodic curves. This enables an approximate comparison to be made between the two speakers regarding the similarity of their pitch/energy contours over the course of an interaction, and is taken to be an indication of effort in prosodic accommodation. The measurement of AUC accommodation for each speaker is derived by dividing the AUC calculation for each speaker by the combined AUC of both speakers and multiplying by 100. The combined AUC accommodation score for both speakers will always total 100 and as such the individual differences can be expressed as a percentage of effort made by each speaker. This process was tested with dummy data to ensure the resultant values accurately reflected the relationship of the speech data. For example, false conversations were created where each side of the conversation was the exact same audio file, resulting in an

exact 50/50 AUC accommodation score in each case.

2.4. CRM evaluation

Two human factors researchers evaluated the pilots' CRM skills. CRM is defined as "the application of human factors in the aviation system" [17]. Thus CRM is concerned with the cognitive and interpersonal skills that pilots need to manage the flight operations on the flight deck. This includes how they manage their resources, i.e. the information that they need to process, the technology that they use to support their tasks and how the pilots delegate tasks to one another in the 'team' on the flight-deck. Pilots use cognitive skills to maintain situational awareness, to solve problems and to make decisions. Pilots interpersonal skills are used to communicate with one another and people "off-aircraft" (i.e. Air Traffic Control, Airline Operations Centre). For this study, the methods used were observations, a modified Social Network Analysis (SNA), Hierarchical Task Analysis (HTA), Process Mapping, coordination Demand Analysis (CDA) and Triangulation (see [15] for more details). In this paper, we present the data of the coordination Demand Analysis.

CRM skills were evaluated in terms of Communication (Comm), Situational Awareness (SA), Decision Making (DM), Mission Analysis (MA), Leadership (Lead), Adaptability (Ad) and Assertiveness (Ass) [18]. All metrics were also gathered into one single mean metric namely Total coordination Mean (TCM). CRM skills were scored for each individual and for the team; for every items discussed and for the whole interaction. Each was rated on a 3-point scale: 1 (low), 2 (mid) and 3 (high). The CRM metrics are in keeping with the JAA and FAA recommendations [19, 15]. The commonality of metrics between the curricula recommendations and the current industry standards for measuring CRM and CDA criteria was considered sufficient for CDA to be justifiable as a means with which to examine CRM.

3. Results

In this section we compare CRM skills, as determined by the human evaluators, with speech measurements automatically extracted from the same recordings. A subset of measures are considered, namely: Comm, TCM, prosodic accommodation, total talk time and total overlap time.

As an illustration of the CRM values obtained from the human evaluators, in Figure 4, the mean TCM values per-item are plotted for each team.

Corresponding team prosodic accommodation values, automatically extracted from the entire 10 min conversations, are provided in Figure 5. Some initial comparisons can be made from these figures: The team with the highest prosodic accommodation value (team 6) in Figure 5, has the highest average TCM value across items in 4, while the team with the lowest accommodation value (team 5) has a flat TCM value across items, just above the mean value of 2 (on a 1–3 rating scale).

To compare these metrics in a more objective way, the correlation between CRM and speech measures was evaluated on a per-speaker, per-item basis. In the CRM reports, several time-stamps were indicated for each discussion item, corresponding to the locations within the recording at which that item was discussed. For each of these time-stamps, a 10 second segment was extracted from the recording. Segments corresponding to the same items were concatenated to create one segment for each of the 15 discussion items in each recording.

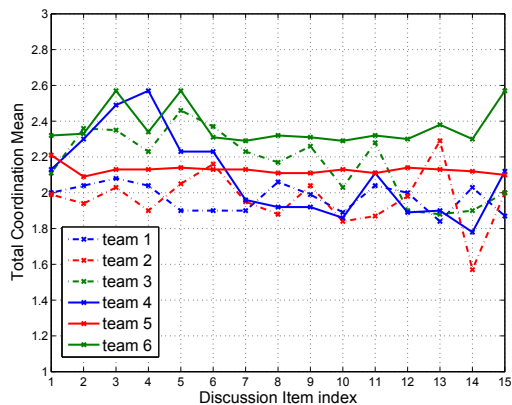


Figure 4: TCM values per-item for each team.

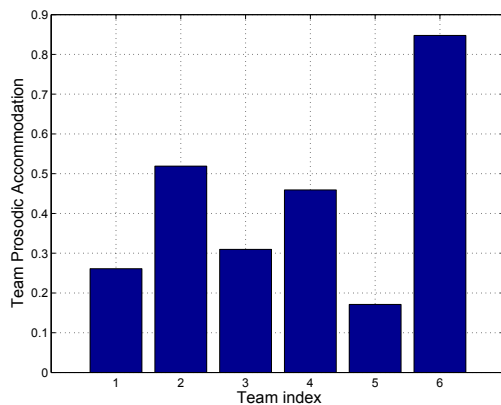


Figure 5: Team Prosodic Accommodation values

For each item segment, automatic speech measures were extracted, and from the CRM reports, corresponding per-item mean values were obtained. The resulting measures were compared by computing the Spearman rank correlation. In Table 1, correlation coefficients for a subset of these comparisons are presented. An issue in extracting the automatic measures was the reliance on a relatively long minimum window length of 30s. After applying VAD, the majority of per-item segments were therefore too short to allow accommodation extraction. As a result, the analysis in Table 1 is based on a reduced sample size of 9 points. By preserving a 30 s window size however, these measures are based on long-term averages (in the case of both the CRM and speech measures), and will be inherently more reliable than short-term measures. Thus, despite the limited sample size, significant correlations were achieved between prosodic accommodation and overlap with both Comm and TCM measurements.

4. Discussion and conclusion

This paper presented a new quantitative speech-based method for objectively assessing communication skills of team members (in this particular case pilots), based on prosodic accom-

	ACC	TLK	OL
Comm	0.73	0.64	0.71
TCM	0.70	0.63	0.75

Table 1: Spearman rank correlation coefficients between Communication (Comm) and Total Coordination mean (TCM) CRM measures, and per-speaker prosodic accommodation (ACC), total talk time (TLK) and overlap time (OL) automatic measures. Statistically significant correlations (in bold) were determined via a two-tailed T-test at a significance level of 0.05.

modation levels and turn-taking organisation. Our results show that the levels of prosodic accommodation and global amount of speaking/ overlap time are indicative of the level of communication and global coordination between members of a team as assessed by human factor experts.

Due to space constraints, we restricted the results presentation to two CRM skills, Comm and TCM, and their correlation to prosodic accommodation and total speaking time and overlaps. Although the analysis was limited by the speech segment durations, significant correlations were observed between automatic and CRM measures. This indicates the potential of the automatic approach to predict CRM measures at the per-item, per-speaker level. The research methodology detailed in the paper continues to serve as a template for continuing the validation work with other pilot groups. To fully evaluate the relative utility of the automatic measures, expanded analyses with more teams assessed, augmented CRM evaluation (using a 5 point likert scale) and an increased number of CRM evaluators have been conducted and further validation is ongoing with preliminary results showing similar correlations and outcomes.

An additional finding of the correlation analysis was that many of the CRM measures are correlated with one another. This is not unexpected, given that the metrics can be subjective and have a limited 1–3 rating scale. This information may be beneficial to CRM evaluators and designers.

Overall, this study has important implications for the assessment of team communication in stressful and safety critical environments. While communication is seen as a non-technical, soft skill, lacking the hard metrics of the more technical skills, it is a vitally important component of any efficient, well-organised team. To this end, adding an objective aspect to the assessment of communication skills to augment current subjective methods, has the potential to improve overall team effectiveness, cohesion and efficiency. This has clear benefits in the area of aviation where effective communication is a necessary skill across the whole aviation domain: pilots, cabin crews, maintenance crews, fire crews; it has benefits in medicine where surgical and nursing teams are increasingly adopting aviation CRM methodology; and it has benefits where any any team of people (from two or more) operate in a complex, safety critical environment.

5. Acknowledgements

This work was undertaken under the PAIA project that was co-funded by the European Regional Development Fund (ERDF) under the National Strategic Reference Framework (NSRF) 2007-2013. We would like to sincerely thank the IALPA pilots who participated in this important study of pilot communication skills via voice analysis and CRM assessments.

6. References

- [1] M. Campion, and E. Papper and G. Medsker, "Relations between work team characteristics and effectiveness: A replication and extension," *Personnel psychology*, vol. 49, no. 2, pp. 429–452. Wiley Online Library. 1996.
- [2] M. Leonard and S. Graham and D. Bonacum, "The human factor: the critical importance of effective teamwork and communication in providing safe care," *Quality and Safety in Health Care*, vol. 13, no. 2, pp. i85–i90. BMJ Publishing Group Ltd. 2004.
- [3] R. Wikander and N. Dahlström, "The Multi Crew Pilot Licence-Revolution, Evolution or not even a Solution?,"
- [4] F. Verheijen, "Flight Training and Pilot Employment," 2002.
- [5] R. Flin and L. Martin and K-M. Goeters and HJ. Hormann and R. Amalberti and C. Valot and H. Nijhuis, "Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills," *Quality and Safety in Health Care*, vol. 3, pp. 97–120. Ashgate. 2003.
- [6] M. A. Rosen, and S. J. Weaver, and E. H. Lazzara, and E. Salas and T. Wu and S. Silvestri, ... and H. B. King, "Tools for evaluating team performance in simulation-based training," *Journal of Emergencies, Trauma and Shock*, vol. 3, no. 4, pp. 353, 2010.
- [7] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception–behavior link and social interaction," *Journal of personality and social psychology*, vol. 76, no. 6, pp. 893. American Psychological Association. 1999.
- [8] M. J. Richardson and K. L. Marsh and R. W. Isenhower and J. R. L. Goodman and R. C. Schmidt, "Rocking together: Dynamics of intentional and unintentional interpersonal coordination," *Human Movement Science*, vol. 26, no. 6, pp. 867–891. Elsevier. 2007.
- [9] K. Shockley and D. C. Richardson and R. Dale, "Conversation and coordinative structures," *Topics in Cognitive Science*, vol. 1, no. 2, pp. 305–319. Wiley Online Library. 2009.
- [10] S. E. Brennan, "Lexical entrainment in spontaneous dialog," *Proceedings of ISSD*, pp. 41–44. 1996.
- [11] H. Giles and N. Coupland and J. Coupland, "Accommodation theory: Communication, context, and consequence," *Contexts of accommodation: Developments in applied sociolinguistics*, pp. 1–68. 1991.
- [12] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *The Behavioral and brain sciences*, vol. 27, no. 2, pp. 169–90. 2004.
- [13] H. P. Branigan and M. J. Pickering and J. Pearson and J. F. McLean, "Linguistic alignment between people and computers," *Journal of Pragmatics*, vol. 42, no. 9, pp. 2355–2368. Elsevier. 2010.
- [14] C. De Looze and B. Vaughan and S. Scherer, "Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction," *Speech Communication*, vol. 58, pp. 11–34. Elsevier. 2014.
- [15] A. M. Kay and P. Liston and S. Cromie, "Measuring Crew Resource Management: Challenges and Recommendations," *Proceedings of Engineering Psychology and Cognitive Ergonomics - 11th International Conference, EPCE 2014, Held as Part of HCI International 2014*, pp. 22–27. Heraklion, Crete, Greece. 2014.
- [16] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, vol. 52, no. 1, pp. 12–40. 2010.
- [17] R. L. Helmreich and H. C. Foushee, "Why CRM? Empirical and Theoretical Bases of Human Factors Training," *Kanki, B., Helmreich, R. and Anca, J. (2010) Crew Resource Management*, 2nd Edition, Elsevier, San Diego, U.S. 2010.
- [18] S. C. Burke, "Team Task Analysis," *Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P. (eds.) Human Factors Methods: A Practical Guide for Engineering and Design*, Ashgate, Hampshire, UK, pp. 56.1– 56.8, 2005.
- [19] R. Flin and P. OConnor and M. Crichton, *Safety at the Sharp End: A guide to non-technical skills*, Ashgate, Surrey, UK. 2008.