Chapter 4
Modelling Conversation

Constable Martin, Dauwels Justin, Dauwels Shoko, Umer Rasheed, Zhou Mengyu and Tahir Yasir

Abstract Conversation is clearly important in our daily lives. Functionally, it serves to deliver and exchange information. However, there is much of a conversation that lays outside of its verbal content, yet impacts directly on those involved and in a manner that might be to their detriment or benefit. For example, in an interview (which is a special class of conversation) the interviewee might needlessly interrupt the interviewer or be too silent, both of which are detrimental to the health of the conversation. This is the non-verbal component of conversation, which is to say it lays outside of the conversation's spoken content. By and large it also lays outside the sphere of what we are consciously aware of. The unsolved problem is how the non-verbal component of a conversation might be visualised in a concise, yet effective manner that would be suitable for use in a communication skill training scenario.

4.1 Learning Conversation Skills

Whether consciously or not, we adjust our voice and body movement when communicating with others. These are skills which we acquire through everyday social engagement. In other words we learn communication skills through experience. AQ1

Experience is a powerful source of learning, especially in the acquisition of soft skills such as human-to-human communication. In such communication, especially

C. Martin (✉) · D. Justin · D. Shoko · U. Rasheed · Z. Mengyu · T. Yasir
BeingThere Centre, Nanyang Technological University, Singapore, Singapore
e-mail: MConstable@ntu.edu.sg

D. Justin
e-mail: JDAUWELS@ntu.edu.sg

D. Shoko
e-mail: SDauwels@ntu.edu.sg

U. Rasheed
e-mail: UMER1@e.ntu.edu.sg

T. Yasir
e-mail: YASIR001@e.ntu.edu.sg

N. Magnenat-Thalmann et al. (eds.), *Context Aware Human-Robot and Human-Agent Interaction*, Human–Computer Interaction Series,
DOI 10.1007/978-3-319-19947-4_4

in that which takes place face-to-face, we focus not only on what we say, but how we say it. The manner in which we speak could change the meaning of what we are saying. For example, if someone smiles and say ‘OK’, it gives a positive impression and most probably means ‘yes’. On the other hand, if they frown, roll their eyes and say ‘Oooo-Kay’, they can come across as displeased. We gather such subtle information about the mental state of others while they are speaking.

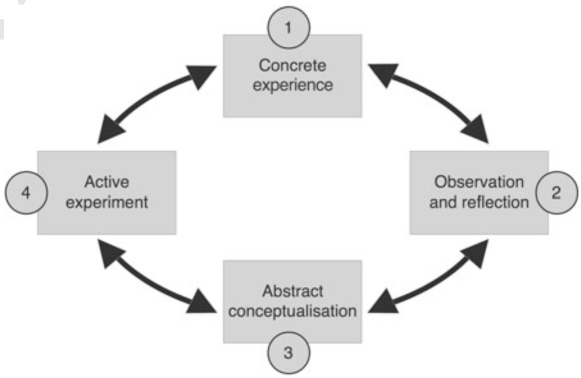
These skills are informed by one’s past experience and may not work as intended in all situations. We learn these skills as children within a small group of culturally homogeneous people, but as we develop and mature it is likely that we will be required to communicate with a far wider variety of people. These people will be of diverse language, culture and religion and will also be diverse in their personality. We, therefore, all need to maintain and upgrade our communication skill set as we grow: to ensure that it is suitable to a wide range of needs.

Since communication skill is practice-oriented, we need real-world experience in order to acquire such skills. However, just by engaging in an activity does not mean that we are necessarily learning from it. How can we knowingly develop such skills? What do we need for such learning/training? We shall discus in this section some of the problems associated with learning within this domain, and we shall propose that a fusion of technology and new generation media provides an effective platform for serving these needs.

Communication skills are quite personal and occur through an invisible cognitive process. Although recent advances in brain science and neuroscience reveal some of the mechanisms underpinning this practice, the brain is still effectively a ‘black box’ and we cannot fully assess how we use non-verbal behaviours while speaking. This type of highly embedded intelligence is known as ‘tacit knowledge’. As tacit knowledge is difficult to articulate, we cannot learn by textbook-based learning. Instead, we need direct experience.

The experiential learning model was proposed by Kolb et al. [11]. This was based on earlier work by scholars engaged in professional learning [3, 15]. In the experiential learning framework (Fig. 4.1), the student follows the four steps of continuing

Fig. 4.1 The experiential learning framework



process: 1 concrete experience, 2 observation and reflection, 3 abstract conceptualisation and 4 active experiment.

Without the students knowing their current performance, it is hard for them to modify their communication behaviour within a conversation. Therein lays the need for feedback in the training process. Such feedback would require that a visualisation of a conversation be available.

4.2 The State of the Art

Every time we look at a traditional internet chat log or an SMS exchange (Fig. 4.2), we are seeing a visualisation of a conversation from which we can derive a significant amount of information. In the latter, the direction from which the speech bubbles originate clearly indicates to whom a remark should be attributed, and the nested response boxes of the former show us the nested sub-topics within a conversation.

Within an SMS exchange there might also be emojis which visualise the emotional subtext of the conversation. Though an emoji is a visualisation of a participant's emotional condition, it is a self-elected one and therefore vulnerable to misrepresentation

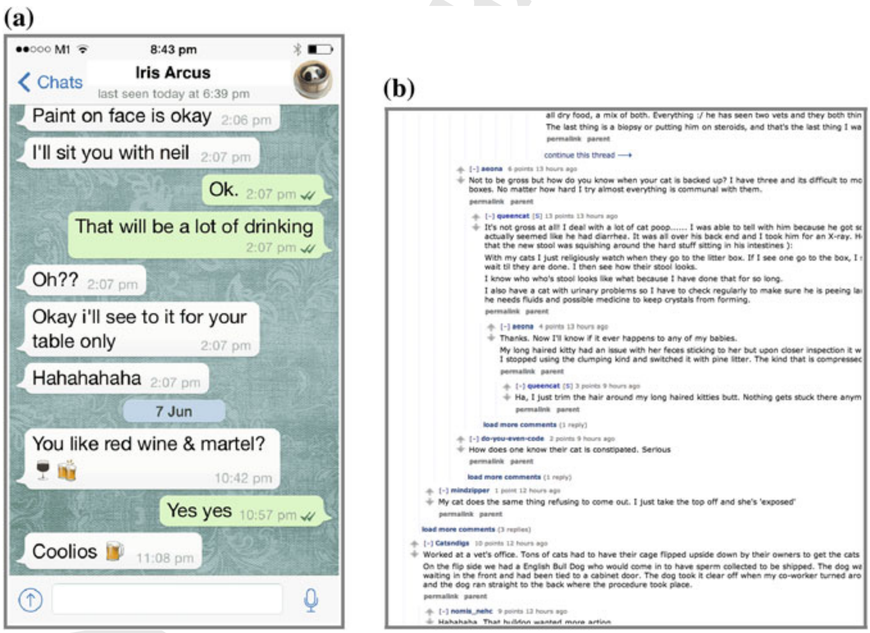


Fig. 4.2 Traditional forms of text-based exchange. **a** An SMS chat exchange showing emojis integrated into the chat. **b** A chat log showing nested conversation

Fig. 4.3 Donath et al's. ‘*Chat Circle*’ interface

and subjectivity. This has not affected their popularity and indeed there now exists chat networks devoted exclusively to emoji [4, 5].

Work by Donath et al. [9] (Fig. 4.3) proposed a multiparticipant text-based chat system (‘Chat Circles’) that added an extra dimension to that which is supported by traditional systems. Each chat participant was represented as a coloured circle. The brightness of each circle indicated the degree of activity of that participant. The proximity of one circle to another was offered to the participant as a dimensional control representing the degree of their engagement with a particular co-participant. As well as offering this extra dimension to a chat conversation, it also visualised social aspects of that conversation.

This and preceding examples, as well as being visualisations of a conversation, are also the conversation itself. There is little distance between the thing and its representation with the latter offering no summary of the former.

A tag cloud, sometimes known as a word cloud (Fig. 4.4), will visualise the frequency of words in a collection of text, with those that have been used most frequently being represented as larger. A degree of summative evaluation may be gathered ‘at a glance’, with important words being signified by their large size.

Tat and Carpendale [23] propose a complex and evaluative approach: ‘Bubba Talk’. This analysed a multiparticipant text-based conversation for such things as the frequency of exclamation marks, the number of words, and the number of characters.

Fig. 4.4 An example of a tag cloud

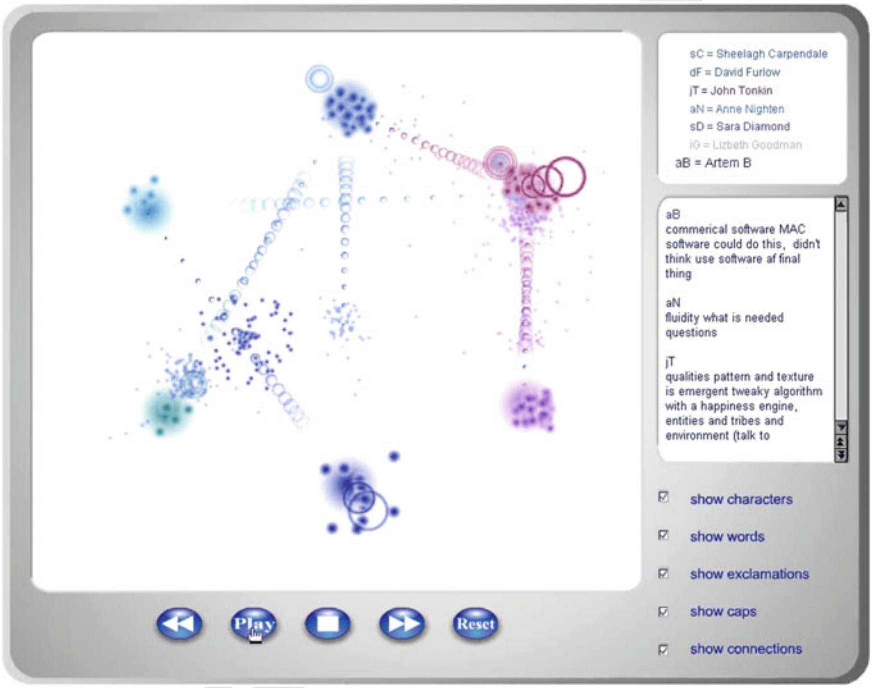


Fig. 4.5 Tat and Carpendale's visualisation of a text chat ('Bubba Talk')

It expressed the entirety of a conversation as a single complex circular-form abstract from which the social dynamics of the conversation may be inferred.

A range of further approaches for visualising text-based chat are reviewed by Uthus and Aha in 'Multiparticipant chat analysis: A survey' [26] ~~(Fig. 4.5).~~ AQ2

Visualising spoken dialogue presents a different challenge to that posed by text, being more dimensionally complex. While communicating with others, humans exchange messages verbally and non-verbally [16]. Non-verbal cues are from communication channels that lay outside of speech (sometimes known as 'paralanguage').

336339_1_En_4_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:22/6/2015 Pages: 32 Layout: T1-Standard

Examples include: eye movement, facial expression, gesture, posture, etc. It is the non-verbal aspects of a conversation that signify its 'health'. For example, what you are saying may be perfectly reasonable and polite, but if you have interrupted someone as you are speaking (which is a non-verbal cue) then you may come across as rude. It should also be noted that although we can easily manipulate what we say in order to create a particular impression, it is far harder to do so using the non-verbal aspects of how we speak. In this sense it is harder to 'lie' non-verbally. However, non-verbal cues do not lend themselves to easy visualisation, and therein lays one of the challenges of our research.

Campbell proposed an approach whereby a spoken conversation was synchronised with a text-based transcription of its content. The length of each spoken utterance was represented by the length of a simple coloured bar on a timeline, the colour of the bar indicated the identity of the speaker. A mouse-over on the bar revealed the transcribed text. Non-verbal cues such as simultaneous speech, interruptions and interjections could be inferred from the relative position of the bars to each other upon the timeline, however this information was not explicitly processed or visualised (Fig. 4.6).

Bergstrom et al. visualised conversations between small groups of people [1]. The output of their approach resembled that of Tat and Carpendale's: a circular-form abstract, this form having a degree of natural suitability to the expression of group conversations. The parameters from which this visualisation was derived were: speaking activity (active/not active) and speaking volume. Secondary (inferred) parameters were: turn-taking and simultaneous speaking. Different to Campbell's approach, theirs did not address the content of the conversation, being instead exclusively concerned with non-verbal cues. Though fascinating, and even beautiful, their approach

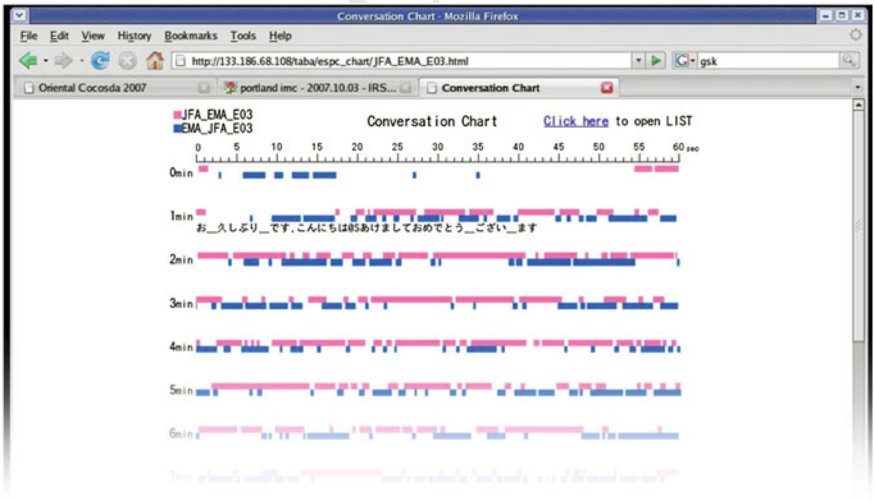


Fig. 4.6 Campbell's transcription interface showing a two-party conversation with the transcribed text (in Japanese) visible in the *top row*

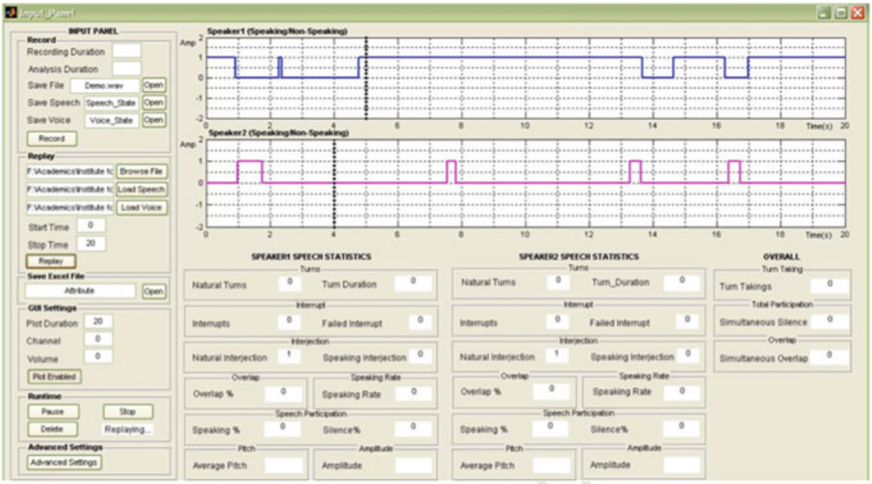


Fig. 4.7 Sarda's et al's. visualisation of non-verbal speech cues

does not on its own offer any high-level evaluation of the health of a conversation. This point will be elaborated upon later on in this chapter in Sect. 4.4.3.

Research in the fields of psychology and cognitive science, in which human behaviour within social interaction is studied, have examined these cues [17, 19] and Gatica-Perez [6] describes ways in which non-verbal cues may be automatically gathered. Referencing such work, Sarda et al. [20] visualised a large number of non-verbal speech cues as a plot along a timeline (Fig. 4.7). This they did using recent advances in recording equipment and signal processing in order to automatically detect these conversation dynamics.

Sarda's approach was not primarily designed for use in a training scenario, being limited for use by researchers wishing to review their data. Additionally, the data it records is low level, being concerned only with statistics, and would therefore have to be interpreted by an expert to have any value in a training scenario.

In summary, a conversation is a dimensionally complex phenomenon involving at least two streams of time-varying data that interact in meaningful and complex ways. There are many ways to visualise a conversation, depending on the form of the conversation and what is required of the visualisation. For the purposes of aiding in the training of conversation skills, we exclusively focused on visualising its non-verbal cues.

4.3 Summary of Our Approach

We focused on cues derived from non-verbal speech. There are several reasons for choosing speech cues as opposed to visual cues. Firstly, speech data can be processed quicker than visual data. In a learning situation, it would be of clear advantage to have feedback that is available on short notice. Secondly, body gesture strongly reflects cultural difference which adds a layer of complexity onto an already complex task. In order to quantify body gesture significant study would be required of its automatic classification and its culturally-specific significance. However, the meaning of non-verbal speech cues is generally universal to all cultures and is therefore easier to classify. One of the leading studies on communication reported that when judging like/dislike, vocal cues were the second most influential channel (38 %) following visual cues (55 %) [14]. For these reasons, speech cues were seen as the best option to produce informative yet speedy feedback.

Our approach was to use technology to detect non-verbal cues and to design a visualisation approach that serves to give feedback to individuals for the purpose of their training. There were four steps to this task:

1. Capturing the non-verbal cues from a conversation (detailed in Sect. 4.4.1)
2. Processing the non-verbal cues as low-level measures (detailed in Sect. 4.4.2)
3. Interpreting the low-level measures as high-level metameasures (detailed in Sect. 4.4.3)
4. Visualising the metameasures for the purposes of training feedback (detailed in Sect. 4.5).

Additionally, the results of a user study are presented in Sect. 4.6.

4.4 The Capture, Processing and Interpreting of Non-verbal Speech Cues

For our purposes the conversation size was restricted to the dyadic. This made our visualisation approach easier to test bed and was also more suitable for a training scenario, which would typically consist of a single trainer/trainee pair. The process required that the raw conversation data was gathered in a manner that did not impact upon its quality. From this data, non-verbal speech cues were automatically gathered and then classified using a number of measures. These measures were the statistical low-level features of the conversation. Using machine learning 3 metameasures were extrapolated from these measures: dominance, interest and discord. These metameasures quantify the high-level 'health' of a conversation.

4.4.1 Protocols for the Capturing of the Speech Data

The following section outlines the step-by-step procedure that constituted our protocol for capturing the speech data from face-to-face dyadic conversations. It was designed to gather data in a controlled manner and without distracting the participants too much.

1. In order to ensure effective communication the recording environment was setup so as to be as non-invasive as possible. Therefore, minimal apparatus was used. For audio recording, we utilised easy-to-use portable equipment for recording conversations. It simply consisted of lapel microphones for each of the two speakers and an audio H4N recorder that allowed multiple microphones to be interfaced with the computer. The speech from each speaker was saved simultaneously in a 2-channel audio *.wav* file.
2. In order to ensure smooth conversation throughout the recording we kept one of the participants constant in the experiment. They acted as a control and facilitated different social scenarios in conversation with their co-participant.
3. In order to obtain a high-quality recording the microphones were attached directly onto the participant's collar. Directional microphones were used so that one speaker's voice did not impede on the other speaker's channel.
4. Both speakers were seated about 1.5 m apart so that each microphone only recorded the voice of the respective participant, and there was no interference from the other participant.
5. The two participants remained in a noise-free environment without any interruptions.
6. The participants were briefed about the experiment and were asked to act naturally. They were also asked to agree on a topic of mutual interest. The topics of discussion ranged from small talk to heated debates on sports, politics, etc. The topics were selected carefully in order to evoke a variety of behaviours.
7. The recording was initiated via a laptop remotely connected to a server.
8. The conversation was monitored remotely via a wireless live feed. Each conversation was around 2.5–3 min in duration and was without any interruptions.

The final speech database consisted of around 100 two-person conversations, each around 1–1.5 min long: a combined total of 200 individual audio recordings. The topics of conversation varied from discussion of assignments, student projects, social and political views etc. The dataset encompassed many distinct social scenarios such as conflicts and disagreements, periods of boredom, aggressive behaviour, storytrading between speakers, speaker-to-speaker exploration, lecturing, etc. This wide range of sociometric samples provided an effective and flexible database.

4.4.2 Processing the Speech Data as Measures

We took from the literature [6] 7 conversational measures (Table 4.1) which together broadly describe the social dynamic of a dyadic conversation and which could also be automatically processed from the speech data.

In Fig. 4.8 the process of deriving a measure from the speech data is visually summarised for the two measures: 'interruption' and 'failed interruption'. The peaks in the plots represent the duration of a participant's speaking. It can be seen that in the second example the speaking duration of speaker *A* lays inside of speaker *B*'s speaking duration. This is classified as a failed interruption.

Table 4.1 The conversation measures derived from the speech data

Measure	Significance
Speaking percentage	The amount of speaking that person *A* or *B* has done expressed as a percentage of the entire conversation
Natural turn taking	The number of times person *A* speaks in the conversation without interrupting person *B*
Turn duration	The average speaker turn duration. The turns of speaker *A* and *B* are both considered
Interjection	The number of times person *A* speaks simultaneous to person *B* but for a period of 1 s or less. This is to indicate short utterances like 'no', 'ok', 'yeah', etc.
Interruption	The number of times person *A*, interrupts person *B* while speaking and takes over the conversation, causing person *B* to stop speaking
Failed interruption	The number of instances when person *A* interrupts person *B* while speaking but stops speaking before person *B* does
Mutual silence	*A* and *B* are both silent

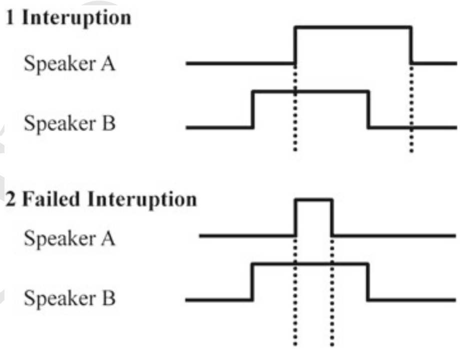


Fig. 4.8 An interruption and a failed interruption evident in the speech data. See how in the second example the speaking duration of speaker *A* lays inside of speaker *B*'s speaking duration. This is classified as a failed interruption

4.4.3 Interpreting the Measures as Metameasures

The measures themselves are statistical low-level features of the conversation and do not by themselves signify any high-level qualitative value. In order for these measures to be of use in a training scenario we summarised them as 3 high-level values: dominance, interest and discord [6]. These are described in Table 4.2. For this a training procedure was developed. This required that a ground truth be established, for which purpose manual classification was required. Each audio recording in the data set was classified manually by at least 5 people. For each recording, they completed a questionnaire relating to their qualitative impression of the speaking mannerisms and behavioural aspects of each participant. The responses range from 1 (low) to 5 (high). For example, if a participant seemed bored, their interest level was classified as being 'low'. In contrast, if they seemed excited, then the interest level was classified as being 'high'. From these five votes the majority view was taken as the final score.

Table 4.2 The three metameasures described in terms of the speech data measures

Metameasure	Significance
Dominance	Dominance indicates the extent of a speaker's influence on their partner as measured by the difference in their **speaking percentage** and the difference in **natural turns**
Interest	Interest indicates the extent of a speaker's engagement with the conversation as measured by the **speaking percentage**, **turn duration** and **interjections**. The more they are involved in the conversation, the stronger the interest they have
Discord	Discord indicates the speaker's lack of agreement with their partner as measured by **interruptions**, **failed interruptions** and **mutual silence**

With the manual classification established as a ground truth, machine learning was applied and we were then in a position where automatic classification could be performed. Using this approach as our basis, a conversation can be automatically classified according to the three metameasures. Each metameasure was expressed in the final output as an intensity value between 0 and 3.

In addition to presenting the complex measures in a summarised and clear form, our approach also normalised the data. For each of the metameasures a long conversation would be subject to the same *n* out of 3 score as a short conversation. The advantage of this is that the length of a conversation is of no significance to the quantification of its quality. This make comparative evaluation of two or more conversations easier to perform.

4.5 The Visualisation of the Data

The task of visualising the data required that its dimensional complexity be recognised. A conversation varies across time and is composed of emotional attributes which are abstract in their nature. Visualising such data is therefore not an easy task.

Additionally, the intended application of our approach is within a training scenario. The exactitude of the visualisation is not as important as its form: it should be clear yet enticing. The metameasures should not just be presented as values but also as *experiences* that the trainee can relate to.

4.5.1 Metaphor and Data Visualisation

The task required that an appropriate model of visualisation be found: one that address the fundamentally abstract nature of non-verbal speech cues.

The heights of a group of people may be visualised as different points on the Y axis within a graph. Here the dots would be operating in a graphical manner and their successful interpretation would depend upon the assumption that the reader is familiar with the convention of how such graphs function. This problem becomes more acute in the case of specialised forms of visualisation such as box plots, histograms and suchlike.

Some things are not suitable to being pictorially visualised in a straightforward manner. For example, how might a volatile political situation be represented? In the preceding examples, there was a clear *indexical* relationship between the heights of the pictograms and the heights of the people. However, given the inherently abstract nature of a political situation, this approach is not feasible. It might be that in such a case a metaphor may be a more effective strategy to employ.

Metaphors rely on our ability to transfer an understanding from one subject to another [12]. In the preceding example, a pictogram of a volcano might effectively signify a volatile political situation. The volcano does not and cannot *visually resemble* a volatile political situation but it is nonetheless possible to read it as such. The disadvantage of a metaphor is that it is inherently ambiguous and therefore its correct interpretation depends upon the reader being privy to the correct way to read it. Thus we find that a metaphoric device such as the inversion of a sign, might variously indicate the opposite of the signified (e.g. an upside down cross signifying satanism), the death of the signified (in *The Book of Signs* Rudolph Koch describes a pictogram of an upside down man as signifying a dead man [10]) or a 'special condition' of the signified (e.g. The figure of the upside down man in Tarot cards can variously mean: acceptance, a new point of view or surrender). We may therefore conclude that a metaphoric visualisation can be subject to multiple interpretations and that context is important in order that a specific reading may be pinned down.

4.5.2 Time and Data Visualisation

A conversation is time-varying in nature. For a human, time is a fundamentally experience-based phenomenon [18] that again presents challenges in its visualisation. Any data that is time-varying requires that time is accommodated as a navigational dimension that is extra to the data. A single value that varies in intensity over time can be presented as a graph on a timeline, as in Sarda et al's. work [20]. However, this is not suitable should the data be more complex such as in the case of several values varying over time. Some existing solutions utilise 3D as this extra navigational space [8, 24] and an example of 3D in everyday use is the depth dimension employed in Apple's Time Machine (their propriety data backup service).

However, what is missed in such approaches is an *experienced* sense of the difference between the beginning and the end. To the user such an experience may allow them to effectively *live* the data and, by proxy, empathise more effectively with the conversation from which the data was derived. We are reminded here that a key need of information visualisation is not just to visualise data but also to communicate effectively, and empathy is a key component of communication.

336339_1_En_4_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:22/6/2015 Pages: 32 Layout: T1-Standard

A possible alternative to a timeline and 3D visualisation is to present the time-varying data as a narrative. There is much previous research on storytelling as an effective means of imparting information and much of it addresses storytelling as an effective way in which to present complex information in a simple and summarised manner [2, 7].

4.5.3 Game Engines and Data Visualisation

It was decided that the most suitable way of presenting the time-varying data was in the form of an animation of two characters engaged in a social exchange in a manner reminiscent of a narrative. Here time was being used to visualise itself, thereby preserving the experience of time, and narrative was being employed to signify change. These characters were interacting with each other, similar to the way that characters interact in games. The form of these interactions was chosen to metaphorically signify the 3 metameasures by which the conversation has been classified.

Visualisations of data can be easily generated using Microsoft's Excel or the open source web app Raw. Using an application like Adobe's Flash or the open source Pure Data it is possible to parse time-varying data into forms which might be animated. However, these approaches are not equal to the task of producing a sophisticated animation. Normally animation, especially that of the human figure, is an arduous task requiring expert input from experienced professionals. This would preclude against their use in a training situation where on-demand feedback would be a key requirement. A simple alternative is to use a game engine. A game engine is a layer of software that supports a digital game. Its job is to manage the physics and appearance of the game world and oversee the rules of the game. It also presents to the game designer the means to author and edit the game.

Game engines have been used before in the visualisation of information [22, 27]. However, the assumption that these approaches make is that the function of a game engine is to make a game. However, game engines have also been used to make stand-alone animations that permit no player interaction. Such animations are commonly known as Machinima, which are hybrids of gaming and film-making. More recently the game engine extension Source Filmmaker [21] has been developed to capture and edit game engine play into the form of an animation for post-capture editing. The advantage of these approaches is the ease with which animations may be made.

Using the Game engine Unity [25] as our development platform we built a visualisation application the purpose of which was to convert the metameasures into a simple animation. Unity was chosen for its flexibility, relative ease of use and the portability of its output.

The animation that a game engine produces is not the same as that which an animator might produce using animation-specific software. It carries with it much of the 'language' of a game: apparent in its loop-form animations, low polygon count figures, sprite overlays (explosions, glows etc.) and simplified camera moves. With these familiar cues come a particular set of expectations from the user: they would be primed to expect from the animation a degree of social engagement that is also likely to directly involve them (i.e. 'gameplay'). This was suitable to the particular demands of our task and provides the contextual underpinning by which the user may make sense of the metameasure metaphor.

4.5.4 Our Approach

The space that the animation is rendered within is of high importance to how the animation will psychologically impact upon the viewer. It was decided that the best option would be to use isometric projection. Different to traditional 3 point perspectival rendering, objects in an isometric projection do not appear larger or smaller according to their distance from the camera. This form of spatial representation is employed in strategy games such as Starcraft and Age of Empires. It is suitable for eliciting in the viewer the 'gods eye' point of view, wherein all characters are of equal importance. This is unlike 3 point perspective that is employed in first person shooters and in which figures which lay nearest the camera are given psychological weightage over those that lay further away. We elected to use isometric projection as we felt was suitable for the purpose of equalising the emphasis given to the two characters/participants.

In the course of the development of visualisation several dead ends were encountered. For example, before the development of the metameasures the collective dynamic of the conversation was expressed using a range of metaphors driven by the low-level measures. A floating platform was employed to reflect the global rate of the 'turn-taking' measure (Fig. 4.9). Should that measure fall below a threshold value (i.e. participants were not equal in the number of times they spoke) then, by the end of the animation the platform would have developed a wobble and the jets holding it up would be emitting black smoke. Here the notion of imbalance served two readings: the literal (the unbalanced state of the platform) and the metaphoric (the unbalanced state of the conversation).

Following the development of the metameasures as a means to summarise the entirety of the conversation, this approach was seen to be extraneous to our needs. Despite this, embodying a sense of collective health using a metaphorical environment remains an enticing idea that we feel is suitable for future exploitation.

We elected to use figures, environments, animations and effects that were similar to those of established gaming traditions. By doing so, we sought to build upon the association of this genre with social engagement and also with the notion of merit acquired through practice (a useful value in training). We purchased these figures, animations and visual effects from commercial resellers of gaming assets and customised them to our needs.

The figures were chosen for their broad similarity to existing 'steampunk' type game characters such as those found in the games Final Fantasy, Sudeki and Kirin. This we felt was suitably outside of any specific worldly context. They were placed

336339_1_En_4_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:22/6/2015 Pages: 32 Layout: T1-Standard

Fig. 4.9 The floating platform as an analogy

within a natural environment which was not so noticeable as to be a distraction, and not so stark as to be disturbing. They were positioned so they were facing each other and were initially animated with a simple loop of an ‘at rest’ motion.

The figures were rigged to respond with pre-defined animations to each of the three metameasures (Table 4.4). ‘Feeding’ the timing of the animations was the metameasure values derived from the conversation data. This was presented in the form of a stream, wherein metameasure ‘events’ were delivered at random intervals. Table 4.3 represents such a stream, the values of which are as follows. Speaker A: Dominance = 3, Interest = 1, Discord 2. Speaker B: Dominance = 2, Interest = 2, Discord 1. Just as there was no one-to-one relationship between the length of the conversation and the length of the animation, as outlined in Sect. 4.4.3, so also there was no one-to-one relationship between the order of these events within the animation and the ordering

Table 4.3 Graphical presentation of an example data stream (key: Dom = dominance, Dis = discord, Int = interest)

Speaker A	Dom	0	Dis	Int	Dom	0	Dis	Dom	0
Speaker B	Dom	0	0	Int	Int	0	Dis	0	Dom

Table 4.4 The metameasures as metaphors

Metameasure	Give animation	Receive sequence	Sprite sequence
Dominance	Figure makes a punching gesture	Figure moves as if electrocuted	Energy ray, emanates from giver and hits the receiver
Interest	Figure makes a wide-arm gesture	Figure twirls	Bubbles and sparkles envelop the receiver
Discord	Figure makes a roaring gesture	Figure places their head in their hands	Rain envelopes the receiver

of the conversation. This served to ensure that the animation did not 'illustrate' the conversation, rather it 'symbolised' it.

The animations were augmented by the use of animated sprites. These sprites were similar in form to those employed in games such as StarCraft, World of Warcraft etc. where they are usually employed to signify such things as spells, explosions and forcefields. The animations and sprites were chosen for their metaphoric similarity to the metameasures. The animations are pictured in Figs. 4.10, 4.11 and 4.12.

The animation was available for viewing almost immediately after the conversation had finished. In a training scenario this is of clear advantage.

The startup screen of the application presented the two participants as two characters: one male the other female. This served to differentiate clearly the two participants. As well as being the point at which the user data was loaded, the users also have the option to swap the gender assignment of their characters. As the training scenario was likely to consist of one trainer and one trainee, it was assumed that only the trainee would be concerned about the gender of their character.

Fig. 4.10 The dominance metameasure animation

B&W IN PRINT

Fig. 4.11 The interest metameasure animation

Fig. 4.12 The discord metameasure animation

The animation in play (Fig. 4.13) presented a running score of the metameasures in the traditional health bar format, which needs no explanation to most people under the age of 50.

As the metameasures did not relate directly to any particular event on the timeline, the animation was effectively functioning as a means by which the metameasure values could be slowly released to the trainee in as much time as the animation lasted (this was set at 1 min and 30 s). This was long enough to serve the purpose of allowing the trainer to discuss with the trainee the metameasures as they arose, yet also was not so long as to risk being tedious to view.

The animation ended with a screen (Fig. 4.14) that summarised the score and gave a brief explanation of what each metameasure signified. The design and function of this followed the format of the traditional statistics screen (a.k.a 'stats') which again is a familiar gaming device.

It was found that a surprising amount of information was available not only from the metameasures themselves, but also from how they combined. For example, high

Fig. 4.13 The animation in play, showing the health bars and a 'dominance' metameasure animation being employed

'interest' and moderate 'dominance' from both participants would indicate that the conversation is going smoothly. High 'dominance' and low 'discord' from one participant would indicate that they might be acting aggressively.

4.6 User Study and Discussion

To test the validity of our approach, we conducted a user study comprised of 4 tasks. 34 students from Nanyang Technological University (Singapore) participated. 17 of these were from the school of Art, Design and Media (ADM), 17 were from the Schools of Computer Engineering and the School of Business. These two groups we term the ADM and non ADM (NADM) groups. 19 were male and 15 were female.

Evaluation of the results of the user study were done by a comparison of two modes of visualisation: our approach and a 2D graphic. The 2D graphic is show in Fig. 4.15. These were also evaluated with respect to the two user groups.

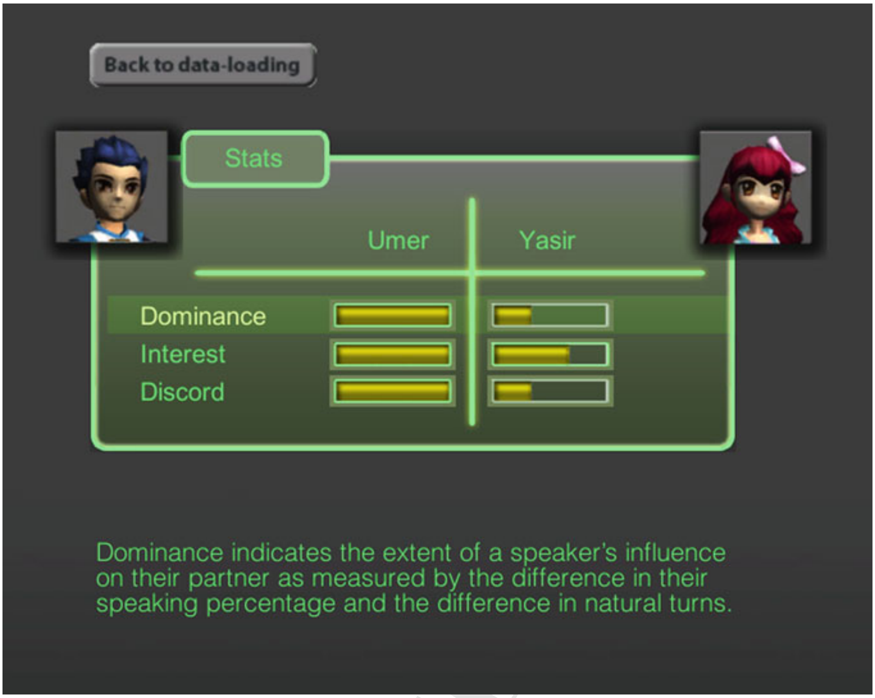


Fig. 4.14 The 'stats' screen, showing the final metameasure score and brief explanations of their significance

4.6.1 Task 1

The participants were shown the 3 animations depicting the 3 metameasures, each 10 s in length. They were then asked to match each animation with its respective metameasure.

The results are summarised in the bar charts Figs. 4.16, 4.17 and 4.18. Given the naturally non-indexical relationship of the metaphor device (i.e. animation) to the signified metameasure, a 100 % success rate in this task was not expected. However, there was nonetheless a high rate of successful pairings for all the metameasures and their respective animations.

Tellingly, the percentage of successful hits for the interest metameasure was slightly higher than that of dominance and discord. This can perhaps be accounted for by the fact that both dominance and discord are emotionally antagonist values and were therefore being confused with each other.

A comparison of bar chart Fig. 4.16 with Fig. 4.17 shows that there was no significant difference between the responses from the ADM and NADM groups.

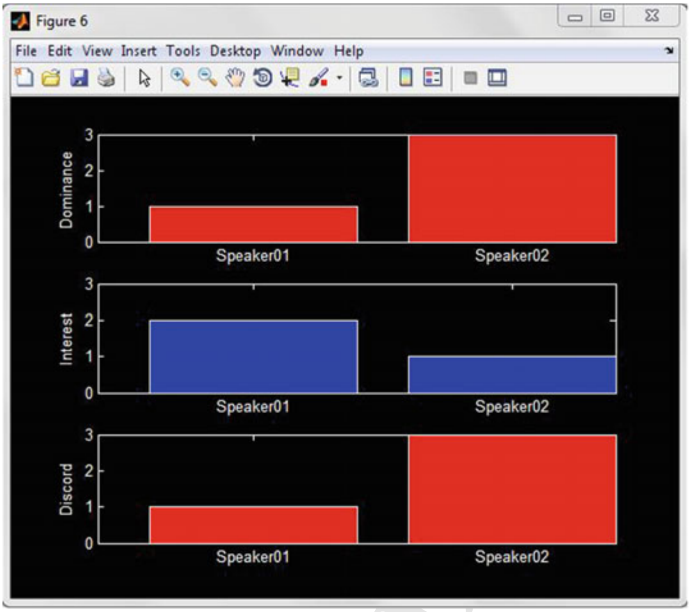


Fig. 4.15 The 2D graphic used in the user study

Fig. 4.16 Results for user study: task 1. *Bars* represent number of successful hits for ADM group

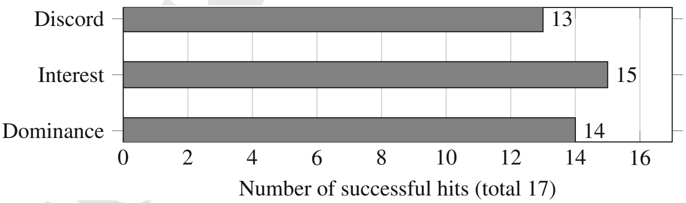


Fig. 4.17 Results for user study: task 1. *Bars* represent number of successful hits for NADM group

336339_1_En_4_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:22/6/2015 Pages: 32 Layout: T1-Standard

Fig. 4.18 Results for user study: task 1. *Bars* represent number of successful hits for ADM and NADM groups combined

4.6.2 Task 2

The participants were asked to listen to three audio recordings of dyadic conversations. Following this they were shown one animation that was generated from the metameasure values of one of these conversations. They were then tasked to match the animation with the correct audio recording.

The success rate for this task correlated strongly with that of task 1, with a correct count of 28/34. This suggests that most of the participants had no trouble extending the principle of the metaphor established in task 1.

Of the 6 incorrect responses, only 2 were from the ADM group. Narrative responses from participants in task 4 (Sect. 4.6.4) provide some illumination as to the reasons for their response. One NADM participant declared that 'as they didn't play video games, the animations were not clear' (response ID 6, Table 4.11). Another from the same group believed that there was a direct one-to-one correlation between the timing of the events in the conversation with the timing of the metameasures animations (response ID 5). This was the only participant to have thought so. One ADM participant talked about the 'character's motivations and intentions' (response ID 13), clearly mistaking the visualisation for a traditional narrative animation.

4.6.3 Task 3

The participants listened to an audio recording of a dyadic conversation. Following this they were shown two visualisations of the metameasures: our animation and a simple graphic in the form of a bar chart. They are then asked to fill in a questionnaire. All responses were tabulated in the Likert style [13]. The questions and their response options are presented in Tables 4.5 and 4.6. The responses themselves are shown in the bar charts: Figs. 4.19, 4.20, 4.21, 4.22, 4.23, 4.24, 4.25, 4.26 and in the Tables 4.7 and 4.8. The results are also broken-down into ADM and NADM responses. Average and standard deviation (SD) values are shown for all sets of results.

Table 4.5 Questions and response options for task 3: animation-related

Response 1	Response 2	Response 3	Response 4	Response 5
Question A: *Was the message clearly conveyed by the animation?*				
Not clear at all	Mostly not clear	Sometimes not clear	Mostly clear	Very clear
Question B: *Did you enjoy the communication training feedback in the form of an animation?*				
Didn't enjoy at all	Mostly didn't enjoy	Neutral feelings	Mostly enjoyed	Very much enjoyed
Question C: *In a communication-training scenario, would you like the feedback to be in the form of an animation?*				
Would not like at all	Mostly would not like	Neutral feelings	Mostly would like	Very much would like
Question D: *Was the length of the animation appropriate to a communication-training situation?*				
Far too short	Too short	Appropriate length	Too long	Far too long
Question E: *Was it helpful that the animation looked similar to a game?*				
Not helpful at all	Mostly not helpful	Neutral feelings	Mostly helpful	Very helpful

Table 4.6 Questions and response options for task 3: graph-related

Response 1	Response 2	Response 3	Response 4	Response 5
Question F: *Was the message clearly conveyed by the graph?*				
Not clear at all	Mostly not clear	Sometimes not clear	Mostly clear	Very clear
Question G: *Did you enjoy the communication training feedback in graphical form?*				
Didn't enjoy at all	Mostly didn't enjoy	Neutral feelings	Mostly enjoyed	Very much enjoyed
Question H: *In a communication-training scenario, would you like the feedback to be in a graphical form?*				
Didn't enjoy at all	Mostly didn't enjoy	Neutral feelings	Mostly enjoyed	Very much enjoyed

Questions A, B and C addressed the participants' response to the animation and were comparable to questions F, G, and H, which addressed the graph. To evaluate the differences between these question pairs, a paired t-test was performed. The results are presented in Table 4.9.

The low t-test result of the question pairs: A/F and B/G indicate that there was significant difference of opinion as to the perceived clarity of the animation and the degree to which it was enjoyed. However, this difference ran in different directions: a majority of participants thought the graph more clear than the animation (question pair: A/F), yet a majority also enjoyed the animation more than the graph (question pair: B/G). The B/G question pair elicited the lowest paired t-test result, indicat-

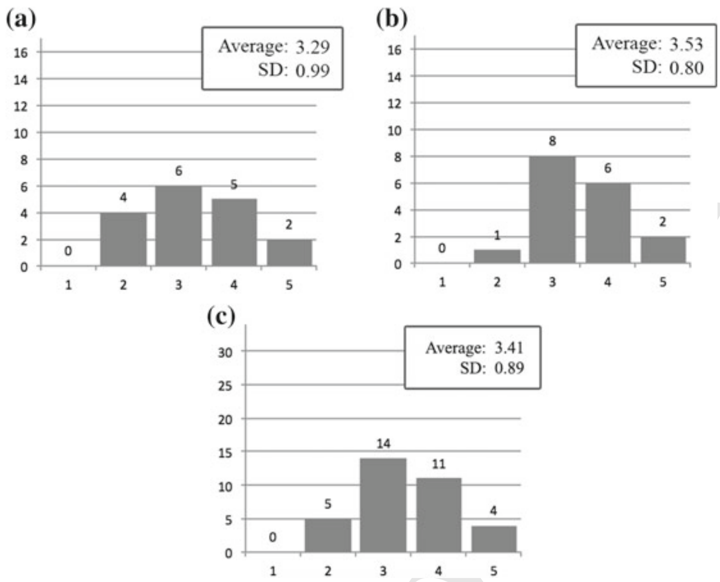


Fig. 4.19 Per-group response to question A: *Was the message clearly conveyed by the animation?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

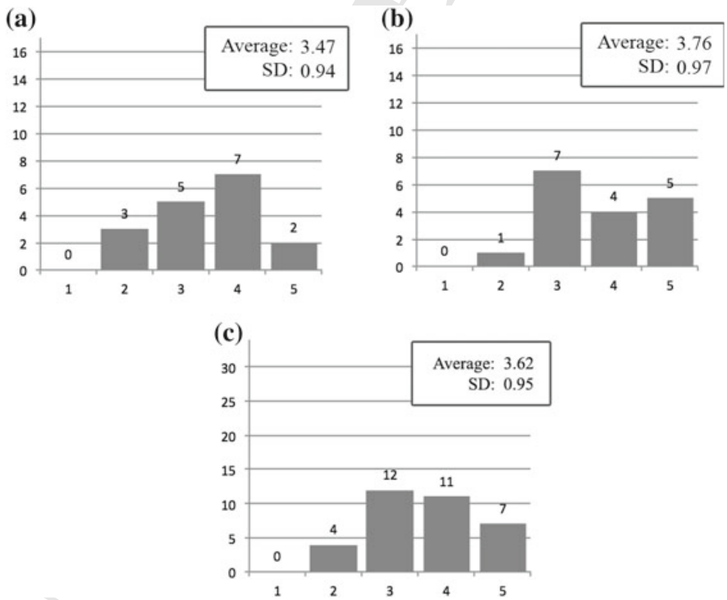


Fig. 4.20 Per-group response to question B: *Did you enjoy the communication training feedback in the form of an animation?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

336339_1_En_4_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:22/6/2015 Pages: 32 Layout: T1-Standard

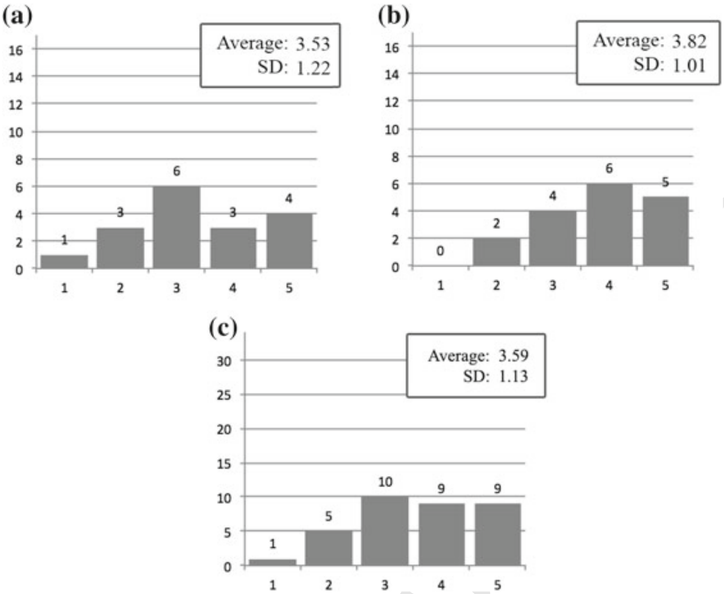


Fig. 4.21 Per-group response to question C: *In a communication-training scenario, would you like the feedback to be in the form of an animation?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

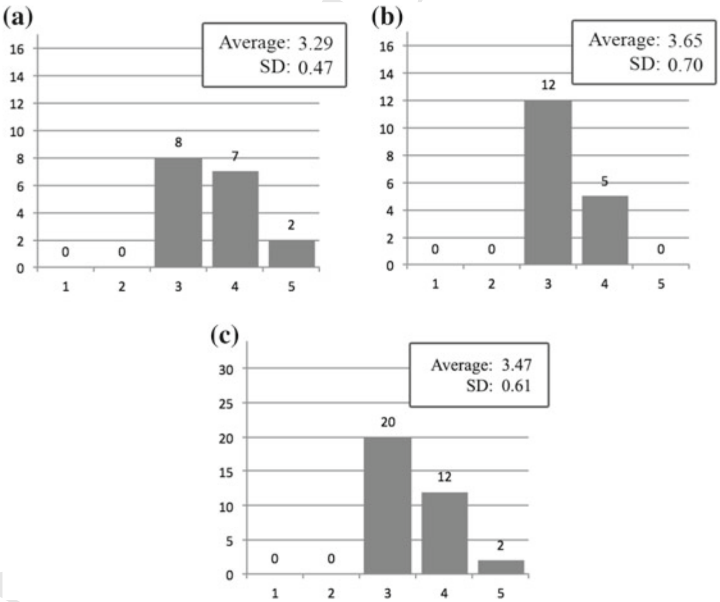


Fig. 4.22 Per-group response to question D: *Was the length of the animation appropriate to a communication-training situation?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

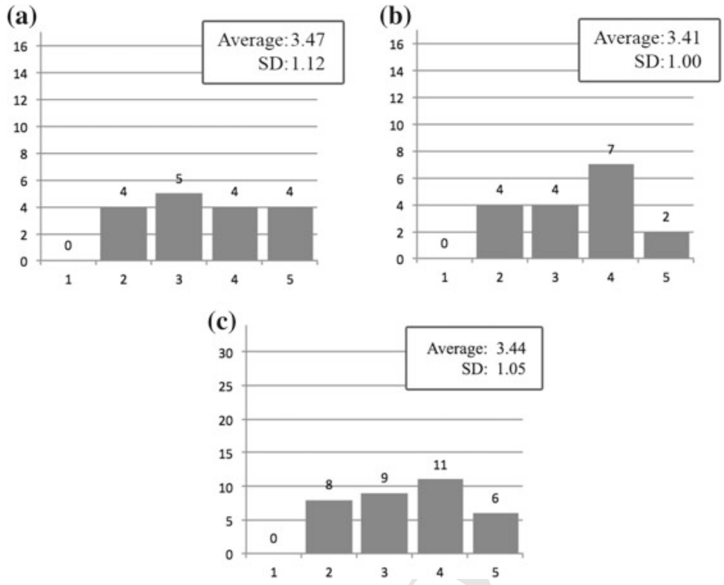


Fig. 4.23 Per-group response to question E: *Was it helpful that the animation looked similar to a game?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

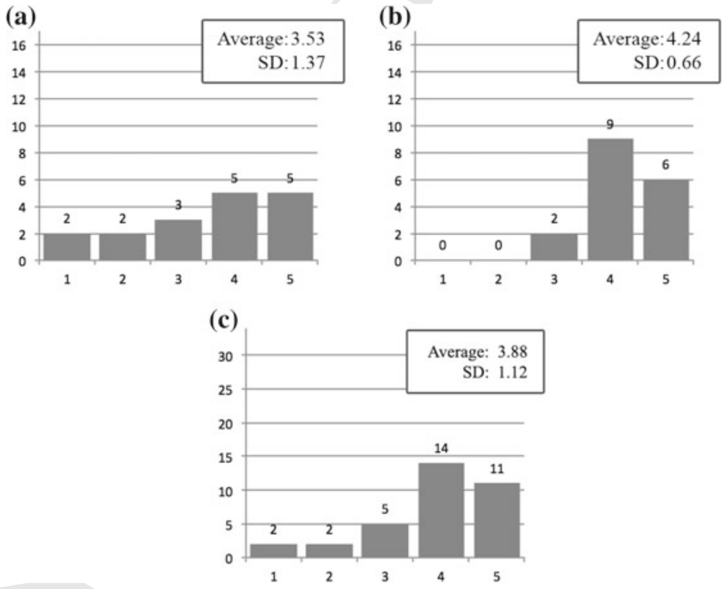


Fig. 4.24 Per-group response to question F: *Was the message clearly conveyed by the graph?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

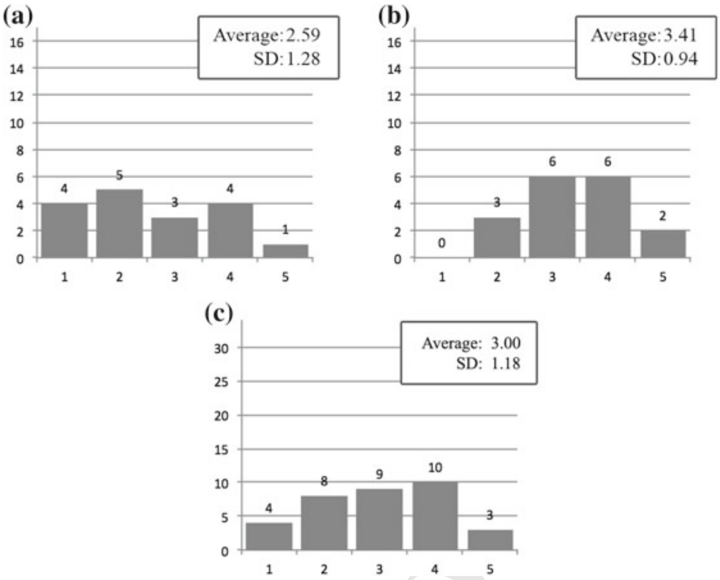


Fig. 4.25 Per-group response to question G: *Did you enjoy the communication training feedback in graphical form?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

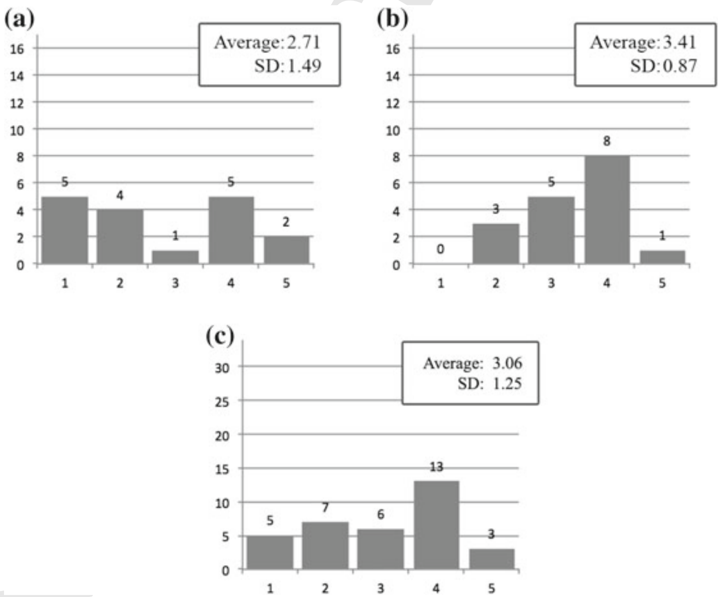


Fig. 4.26 Per-group response to question H: *In a communication-training scenario, would you like the feedback to be in a graphical form?* **a** Response from ADM group. **b** Response from NADM group. **c** Response from all participants

Table 4.7 Responses for task 3: animation-related

	Response count					Response percent				
	1	2	3	4	5	1	2	3	4	5
A	0	5	14	11	4	0	15	41	32	12
B	0	4	12	11	7	0	12	35	32	20
C	1	5	10	9	9	3	15	30	26	26
D	0	0	20	12	2	0	24	26	32	18
E	0	8	9	11	6	0	24	26	32	18

Table 4.8 Responses for task 3: graph-related

	Response count					Response percent				
	1	2	3	4	5	1	2	3	4	5
F	0	4	7	7	4	6	6	15	41	32
G	4	8	9	10	3	15	23	26	29	9
H	5	7	6	13	3	15	20	18	38	9

Table 4.9 Paired *t*-test results of question pairs (* significant at 0.05 level and below)

Question	Average	SD	Paired t-test result
A	3.41	0.89	0.05*
F	3.88	1.12	
B	3.62	0.95	0.03*
G	3.00	1.18	
C	3.59	1.13	0.11
H	3.06	1.25	

ing strong difference with the majority favouring the animated feedback. Narrative responses in Sect. 4.6.4 shed some light on this, with some referring to its 'cuteness' and how it 'sparked their imagination'.

When asked which form of visualisation was more clear, the participants favoured the graph, though the near-borderline t-test result indicated that the difference was not extreme. This favouring was not a surprise, as the aim of our approach was never to present information in as clear a manner as possible, but in a form that was agreeable to the user and suitable for the needs of the training scenario.

There was no significant difference of opinion evident in the C/H question pair, indicating that participants were equally divided as to whether they favour the animation or graph being used in a training scenario.

To test the difference of response between the ADM and NADM groups, a non-paired t-test was performed on the per-group responses. The results are shown in Table 4.10. The low non-paired t-test results of per-group question responses for F and G showed significant per-group difference in the perceived clarity (F) and

Table 4.10 Per-group non-paired t-test results (* significant at 0.05 level and below)

Question	Average	SD	Non-paired t-test result
A ADM	3.29	0.99	0.45
A NADM	3.53	0.80	
B ADM	3.47	0.94	0.37
B NADM	3.76	0.97	
C ADM	3.53	1.22	0.23
C NADM	3.82	1.01	
D ADM	3.29	0.47	0.09
D NADM	3.65	0.70	
E ADM	3.47	1.12	0.87
E NADM	3.41	1.00	
F ADM	3.53	1.37	0.06*
F NADM	4.24	0.66	
G ADM	2.59	1.28	0.04*
G NADM	4.41	0.94	
H ADM	2.71	1.49	0.10
H NADM	3.41	0.87	

enjoyment (G) of the graph, with the ADM group more likely to favour the animation (although this difference was borderline in the case of its clarity). The low SD of the NADM group in response to question F indicated broad agreement of opinion, with most declaring the graph to be clear. However, amongst the ADM group the SD value was quiet high, indicating general disagreement.

Generally, the SD value of the ADM group in answer to all questions, was higher than that of the NADM group, indicating less general agreement than the NADM group. This perhaps can be accounted for by the fact that visual art attracts both very technical students (as in the case of animation) and very visual ones (as in the case of graphics). The nature of their diversity of interests is likely to influence the form of visualisation that the participants favour.

Considering the natural interest and skill domains of these two groups of students, the differences in their responses comes as no surprise. However, the low non-paired t-test value of the animation-related questions A to E indicates that both groups of participants were in broad agreement as to its value.

Which discipline the participants majored in should not be assumed to be the only factor at play in influencing their responses. How familiar they were with the gaming oeuvre would certainly have impacted on their ability to successfully interpret the results. This is borne out by the narrative responses presented in Sect. 4.6.4, particularly response ID 1 and 6 (NADM and ADM participants respectfully). A few of these responses we have correlated to those of task 2, Sect. 4.6.2.

4.6.4 Task 4

The final task was of an open variety: inviting the participant to comment on any aspect of the user study. Most of the responses were perfunctory: reiterating preferences already stated in task 3. However some were more informative and gave us unique information. Notable responses are given verbatim in Table 4.11. Some of these responses are discussed in the preceding sections.

Predictably, the ADM group were inclined to make suggestions as to how the creative aspects of the approach might be improved. Response ID 1 and ID 6 indicate correspondence between a participant's familiarity with gaming and their ability to interpret the animation successfully. Of the 4 from the NADM group that gave an incorrect response to task 2, 1 thought that there was a one-to-one correspondence

Table 4.11 Verbatim responses to user study: task 4

ID	Group	Response
1	NADM	Maybe you can define more choices like dominance, interest and discord. One scenario can be described by multiple tags. (note: this participant remarked verbally that he found the animation easy to read as he was an avid game player)
2	NADM	Characters should be same gender
3	NADM	The people in the animation are cute
4	NADM	The animation was vivid and sparked my imagination
5	NADM	Interesting survey. The first part was a bit confusing and some clips could be categorised in two categories. In Task 3 I tried to match the activity in animation with that in the conversation
6	ADM	As I don't play video games, animation does not work well for me. Instead, I feel graph is easier and direct to understand
7	ADM	There should be music for each animation in Task 1 also that will make is easier to match
8	ADM	I think there should be a balance between animation and graphical summary
9	ADM	Message is harder to convey using animation perhaps game characters are too distractive, simple and straight forward animated expression might help. Message from graph is clearer but less interesting than animation
10	ADM	The animation could include facial expressions to better express the character's feelings. The poses are also a little too subtle, can be made more dynamic for clarity
11	ADM	The background sound during the animation does not really suit well
12	ADM	Facial expressions on the animated characters will be even more helpful in explaining the sociometrics
13	ADM	While the content of the animation is clear on its own, they left me confused when they are played back to back, leaving me scratching my thoughts on the character's motives and intentions

between the animation and the conversation (ID 5). This was the only participant to have done so. This shows that the timing strategy of the metameasure animations, as outlined in Sect. 4.5.4, presented no problem for the majority of respondees.

4.7 Conclusion

An approach was developed that could deliver an animated metaphoric visualisation of the salient non-verbal speech cues of a dyadic conversion. We believe that it could serve as a suitable framework for the delivery of training feedback in a communication skill training scenario. From the analysis of the user studies we may conclude that the goals of our project were satisfactorily achieved.

Our approach was never intended to be better than a simple graphical approach as a means of precisely presenting information, however the results show that it nonetheless presents information in a manner that is clear enough for the stated purposes: to serve as a means by which a trainer may deliver salient feedback as to a trainee's conversational skills. Where it excels is presenting the information in a manner that the trainee could enjoy and could experientially relate to.

Some user study participants made suggestions as to how our approach may be improved. These might be incorporated in further work.

In the selection of the animations and sprites it was required that there be a metaphoric correspondence between them and the metameasures. They were chosen by the authors using their experience in animation and not by any exact empirical method. Exactly by what terms this correspondence exists is a topic into which we did not delve in detail. It encompasses such diverse disciplines as: cognitive linguistics, perception and neurology. Should our approach be expanded it is suspected that a more comprehensive involvement of such disciplines would be required.

References

1. Bergstrom T, Karahalios K (2007) Conversation clock: visualizing audio patterns in co-located groups. In: 2007 40th annual Hawaii international conference on system sciences (HICSS'07), pp 78–78
2. Denning S (2001) The springboard: how storytelling ignites action in knowledge-era organizations. Routledge
3. Dewey J (1962) The relation of theory to practice in education. University of Chicago, Chicago
4. Emojicate (2014) http://emojicate.com/. Accessed 29 Oct 2014
5. Emojli (2014) http://emoj.li/. Accessed 29 Oct 2014
6. Gatica-Perez D (2008) Automatic nonverbal analysis of social interaction in small groups: a review. Image Vis Comput 27(12):1775–1787
7. Gershon N, Page W (2001) What storytelling can do for information visualization. Commun ACM 44(8):31–37
8. Johansson J (2008) Efficient information visualization of multivariate and time-varying data
9. Judith D, Karrie K, Fernanda V (1999) Visualizing conversation. J Comput-Mediat Commun 4(4)

10. Koch R (2013) The book of signs. Courier Dover Publications, Dover
11. Kolb DA et al (1984) Experiential learning: experience as the source of learning and development, vol 1. Prentice-Hall, Englewood Cliffs
12. Kovecses Z (2002) Metaphor: a practical introduction. Oxford University Press, Oxford
13. Likert R (1932) A technique for the measurement of attitudes. Archives of psychology
14. Mehrabian A, Ferris SR (1967) Inference of attitudes from nonverbal communication in two channels. J Consult Psychol 31(3):248
15. Mezirow J (1997) Transformative learning: theory to practice. New Dir Adult Continuing Educ (74):5–12
16. Pentland AS (2010) Honest signals. MIT press, Cambridge
17. Poole MS, Hollingshead AB, McGrath JE, Moreland RL, Rohrbaugh J (2004) Interdisciplinary perspectives on small groups. Small Group Res 35(1):3–16
18. Pöppel E (1978) Time perception. In: Handbook of experimental psychology. Springer, London, pp 713–729
19. Salas E, Sims DE, Burke CS (2005) Is there a "big five" in teamwork? Small Group Res 36(5):555–599
20. Sarda S, Constable M, Dauwels J, Dauwels S, Elgendi M, Mengyu Z, Rasheed U, Tahir Y, Thalmann D, Magnenat-Thalmann N (2014) Real-time feedback system for monitoring and facilitating discussions. In: Natural interaction with robots, knowbots and smartphones. Springer, London, pp 375–387
21. Source filmmaker (2014) http://www.sourcefilmmaker.com. Accessed 29 Oct 2014
22. Stock C, Bishop ID, O'Connor A (2005) Generating virtual environments by linking spatial data processing with a gaming engine. In: Proceedings 6th international conference for information technologies in landscape architecture
23. Tat A, Carpendale MST (2002) Visualising human dialog. In: Sixth international conference on information visualisation, pp 16–21
24. Tominski C, Schulze-Wollgast P, Schumann H (2005) 3d information visualization for time dependent data on maps. In: Information visualisation, 2005. Proceedings, IEEE, pp 175–181
25. Unity (2014) https://unity3d.com. Accessed 29 Oct 2014
26. Uthus DC, Aha DW (2013) Multiparticipant chat analysis: a survey. Artif Intell 199:106–121
27. Wünsche BC, Kot B, Gits A, Amor R, Hosking J (2005) A framework for game engine based visualisations. In: Proceedings of image and vision computing. Citeseer, New Zealand