Quantifying Conversational Facial Gestures

Ella Cullen, Patrick Healey and Paraskevi Argyriou

August 1, 2023

Quantifying Conversational Facial Gestures

Ella Cullen[1], Patrick Healey[1] and Paraskevi Argyriou[2]

[1] School of Electronic Engineering and Computer Science, Queen Mary University

[2] School of Biological and Behavioural Sciences, Queen Mary University

Author Note

We thank Simone Swaby for her assistance with data collection. Correspondence should be addressed to phd candidate researcher Ella Cullen, Queen Mary University, email: e.l.a.cullen@qmul.ac.uk.

Abstract

Conversational facial expressions- ‘puzzled faces’, ‘thinking faces’ and brief eyebrow raises- play an important role in building shared understanding. However, we know much less about these expressions than facial displays of emotion. The existing literature classifying conversational facial expressions has relied on manual coding of facial action units and performed conversational facial expressions. Computer vision techniques make it possible to automatically capture the structure and dynamics of conversational facial expressions across large datasets. This paper demonstrates how this method can be used to capture quantitative measures of facial action units intensities in the ‘puzzled face’ in a corpus of spontaneously-produced conversational facial expressions.

Keywords: Conversational facial expressions, pragmatics, dialogue, multimodal pragmatics, quantitative methods

1. Quantifying Facial Gestures

The majority of work on facial gestures has focused on expressions of emotion. Conversational facial expressions have received much less attention in the literature, despite playing a significant role in the basis of communication; reaching and maintaining mutual understanding between interlocutors (Homke et al., 2022; Nota et al., 2021). Existing classification research of conversational facial expressions has used manual coding to judge whether a display of various facial action units is absent or present (Ekman & & Friesen, 1978; Homke et al., 2022). The facial action unit system (Ekman & Friesen, 1978) is an anatomically-based categorised system comprising of dynamic movements of individual facial action units, such as ‘eyebrow raise’ or ‘chin raise’. This qualitative method of characterising and coding facial gestures shows high inter-rater reliability (Ekman & Friesen, 1978), allowing for standardised categorisation and hence investigation of facial expression components across the literature. However, the threshold judgement of present/absent leaves missing information about the dynamism and intensity of facial action units over time, which is an important factor for the recognisability of facial expressions (Bevacqua et al., 2007). The data curves provided from quantitative corpus data on naturally-displayed facial gestures will allow for more precise characterisation of conversational facial expressions through clusters of intensities of facial action units and provide insight into the interpolation of facial action units over time, allowing for the dynamism of the face to be included in analysis.

One conversational facial expression in particular that provides a good test case of this approach is the ‘puzzled face’. The manual coding of the puzzled face in the literature has relatively consistently identified the key defining facial action unit of a puzzled face to be the eyebrow furrow (Bavelas & Gerwing, 2011; Chovil, 1991; Homke et al., 2022). However, there is a lack of agreement on exactly which other action unit components are involved in the puzzled face (Cunningham et al., 2005; Ekman & Friesen, 1978). The ‘thinking face’ proposed here has not been documented in much literature. However, it was noticed during piloting of various dialogue tasks (10 dyads). This facial display contains similar action units to the puzzled face, particularly the eyebrow furrow. However, this display demonstrates other facial action unit differences and seeming functionality in conversation. The divergence tended to be that the face pointed away with eyes closed, and usually occurred with request to have the floor to allow for a pause while they think, a contrast to the proposed repair-initiation function of the puzzled face (Homke et al., 2022).

As conversational facial expressions are beginning to gain research attention, this paper presents a methodology for more fine-grained analysis that bridges the gaps that previous qualitative, threshold coding introduces. This study takes advantage of advances in computer vision as a viable technique for capturing and profiling conversational facial expressions. A corpus of spontaneously-produced puzzled face displays is analysed to test the viability of automatic capture in live dialogue rather than analysis of performed conversational facial expressions (Kaulard et al., 2012; Guerdelli et al., 2022), that may differ from natural expressions particularly through more stereotyped movement. Therefore, this study’s design should allow for more accurate facial expression analysis with the research question: “Can spontaneously-occurring puzzled face displays be quantitatively characterised within- and between-subjects through cluster analysis of facial action unit intensities?”.

2. Methods

Participants

46 individuals participated in this corpus study in return for a £10 Amazon gift voucher. The inclusion criteria was to be a first-language speaker of English and to be over 18. The participants formed 23 dyads (8 female-male, 7 male-male, 8 female-female). The mean familiarity between interlocutors was 8.56. Therefore, pairs were mainly familiar with each other with a friendly relationship. The mean age of participants was 20.86 years.

Materials

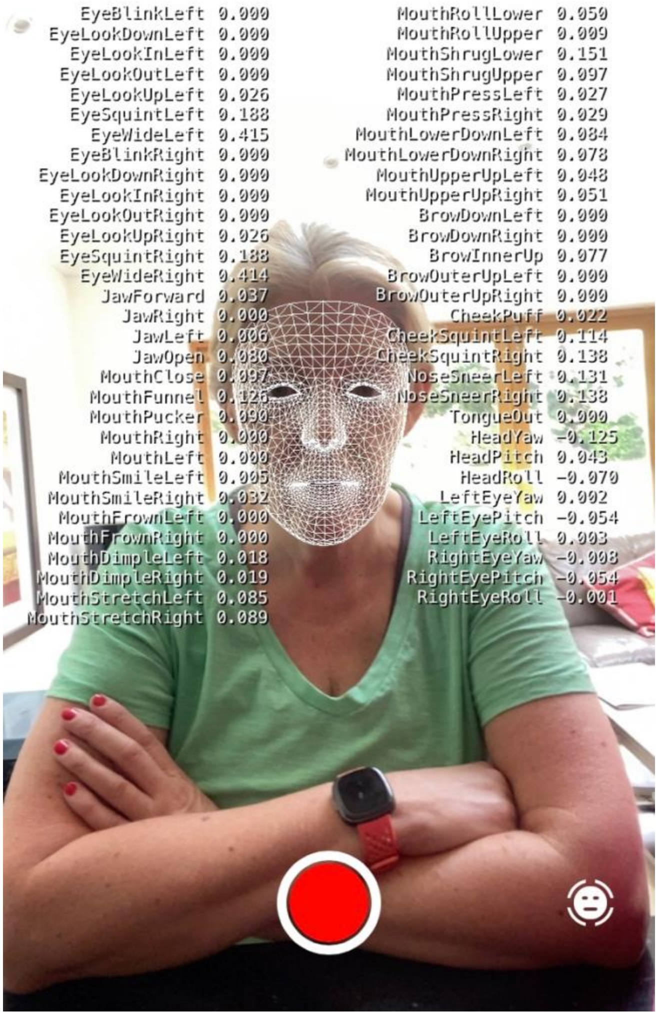
A task was developed to elicit puzzled face displays through a guessing game that manipulates levels of difficulty in the materials participants are asked to describe. A task was developed in which a describer had to describe either abstract concepts (e.g. "To deal with a problem", "To intervene") (Chu et al., 2014) or famous figures and places (e.g. "Martin Luther King Jr.", "Morocco") (Moor et al., 2017; Skowronek & Schoenenberg, 2017) as they appeared on the deck of cards in front of them. The guesser would then have to guess what the card said from the describer's description. However, it was made clear to the guessers that this was an 'open dialogue' and that they could ask probing questions and make incorrect guesses until they reached the solution. Roles switched halfway. The participant on the left always had the first deck of cards, thus beginning roles were randomised rather than chosen.

Equipment

Two iPhones running the Live Link Face App (Epic Games, 2023) were set up facing each participant. The Live Link Face App captures facial movement through the TrueDepth camera technology in the iPhone. Figure 1 illustrates the app set up for facial capture.

Figure 1.

Screenshot of the Live Link Face app's facial motion capture set up.



Procedure

Both participants sat across from one another. An iPhone was set in front of each participant at chin-level (mounted on tripod). There were no limits on how long the dyads spent on each trial, just that they should not move onto the next card until the guesser either successfully identified the target item or the experimenter allowed the dyad to move on. This was to ensure participants were motivated to solve misunderstandings and thus elicit pragmatic facial displays, rather than move on as soon as there was a problem in understanding. The describer picked up one card at a time. Once the describer had gone through all their cards, the roles switched and the other participant went through describing their deck of cards. After the task, participants were fully debriefed.

Data analysis

The first step of data analysis involved identifying intuitively-recognisable displays of the puzzled face. Firstly, a coding manual was developed through the literature's description of the puzzled face and the observational data of this study. 10 naïve coders were then recruited to code a random subset of the videos in order to provide inter-coder reliability with the main coder of this study. Through ELAN software (Wittenburg et al., 2006) and following the coding manual provided to them, their task was to annotate sections of the video in which they recognised the display of a puzzled face. With each annotation of a puzzled face, they were also instructed to rate their level of confidence in their judgement that a puzzled face was being displayed from 1-10. The author (main coder) coded all the 46 videos for displays of the puzzled face, following the same coding manual provided to the naïve coders.

3. Preliminary results

3.1 Inter-rater reliability

Coding from naïve coders has not yet been completed. Therefore an inter-rater reliability score cannot be provided currently.

3.2 Descriptives

Instances of puzzled face (out of 20 coded videos): 117

Mean duration: 2.9 seconds

Mean occurrence in task dialogue: 8

3.3 Mean facial action unit intensities of the puzzled face at peak frame

Table 2.

Mean intensity of key facial action units at peak frame of puzzled face expression

Facial action unit	Peak frame mean intensity
Brow down right	.19
Brow down left	.19
Nose sneer right	.15
Nose sneer left	.14
Cheek squint right	.14
Cheek squint left	.13
Eye squint right	.16
Eye squint left	.16

3.4 Classification of the puzzled face through cluster analysis of mean action unit intensities

Table 1.

Facial action units with the highest intensities throughout the display of the puzzled face for both clusters

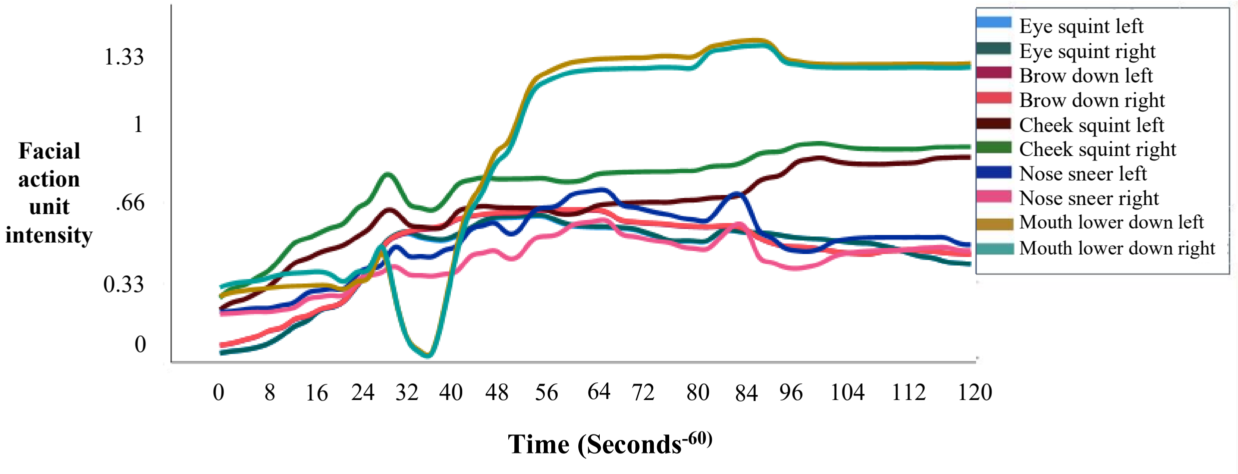
Facial action unit	Mean intensity cluster 1	Mean intensity cluster 2	ANOVA sig.
Brow down right	.22	.16	.097
Brow down left	.22	.16	.099
Nose sneer right	.35	.11	<.001
Nose sneer left	.34	.11	<.001
Cheek squint right	.31	.13	<.001
Cheek squint left	.28	.12	<.001
Eye squint right	.21	.17	.332
Eye squint left	.22	.17	.327

3.3 Interpolation of the most representative puzzled face

3.4

Figure 2.

Line graph to illustrate the interpolation of key action units intensities between the onset and offset of a puzzled face display



Note. This highlights that eyebrows move down, cheeks squint, nose sneers and eyes squint at similar rates of interpolation at the beginning 24 seconds^{-60} of the sequence up to peak frame (32 seconds^{-60}). Then, these action units are slowly disengaged throughout the rest of the display to offset except for the mouth lowering down at increasing intensity and the cheek squint continuing.

Discussion

These preliminary findings demonstrate the viability of computer vision to profile conversational facial expressions, and dialogue task design to elicit spontaneously-produced conversational facial displays for measurement. The data suggest that the key components of the puzzled face are the following facial action units: eyebrow furrow ("brow down"), nose sneer, cheek squint and eye squint. These key facial action units were found to be consistent between the two clusters, cluster 1 was just more intense overall (See Table 1). This finding suggests qualitative consistency in the display of the puzzled face (so between-subject consistency in the facial action units employed during the display of a puzzled face), but that the display can vary along a continuum of intensity.

To graphically demonstrate the interpolation of a characteristic puzzled face, the most representative puzzled face of cluster 1 was used to demonstrate the interpolation of facial action units (see Figure 2). This is due to both clusters sharing the same key action units, but with cluster 1 being more intense. As the naïve coding phase is only currently being carried out and is incomplete, confidence ratings and inter-coder reliability have not yet been obtained so these preliminary findings have attempted to use the cluster analysis to estimate the most intuitively-recognisable puzzled faces. The most intense puzzled face is more likely to be the most recognisable to non-experts and so most likely to be coded by the naïve coders (Kaulard et al., 2012). Hence, cluster 1 was chosen for this illustration of the interpolation of facial action units during a display.

Analysis of the role of conversational facial expressions in the pragmatics of dialogue has been relatively scarce in the dialogue literature. Analysis such as this could allow for multimodal dialogue act tagging as more standard measures of conversational expressions during conversation can be measured. It is clear from the sheer instances of the puzzled face in this study that there is indeed much visual feedback during dialogue.

References

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening*, *25*(3), 178-198. https://doi.org/10.1080/10904018.2010.508675

Bevacqua, E., Mancini, M., Niewiadomski, R., & Pelachaud, C. (2007). An expressive ECA showing complex emotions. In *Proceedings of the AISB annual convention, Newcastle, UK*(pp. 208-216).

Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Research on Language & Social Interaction*, *25*(1-4), 163-194. https://doi.org/10.1080/08351819109389361

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy.*Journal of Experimental Psychology: General*, *143*(2), 694–709. https://doi.org/10.1037/a0033861

Cunningham, D. W., Kleiner, M., Wallraven, C., & Bülthoff, H. H. (2005). Manipulating video sequences to determine the components of conversational facial expressions. *ACM Transactions on Applied Perception*, *2*(3), 251-269. https://doi.org/10.1145/1077399.1077404

Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*. https://doi.org/10.1037/t27734-000

Epic Games (2023, January 21) *Recording Facial Animation from an iOS Device.* Unreal Engine. https://docs.unrealengine.com/5.0/en-US/recording-face-animation-on-ios-device-in-unreal-engine/

Guerdelli, H., Ferrari, C., Barhoumi, W., Ghazouani, H., & Berretti, S. (2022). Macro- and Micro-Expressions Facial Datasets: A Survey. *Sensors (Basel, Switzerland)*, *22*(4), 1524. https://doi.org/10.3390/s22041524

Hömke, P., Levinson, S. C., & Holler, J. (2022). Eyebrow movements as signals of communicative problems in human face-to-face interaction.

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., & Wallraven, C. (2012). The MPI facial expression database--a validated database of emotional and conversational facial expressions. *PloS one*, *7*(3), e32321. https://doi.org/10.1371/journal.pone.0032321

Moor, K.D., Arndt, S., Ammar, D., Voigt-Antons, J., Perkis, A., & Heegaard, P.E. (2017). Exploring diverse measures for evaluating QoE in the context of WebRTC. *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 1-3.

Nota, N., Trujillo, J. P., & Holler, J. (2021). Facial signals and social actions in multimodal face-to-face interaction. *Brain Sciences, 11*(8), 1017. https://doi.org/10.3390/brainsci11081017

Skowronek, Janto, & Schoenenberg, Katrin. (2017). 3CNG – 3-party Celebrity Name Guessing task for conferencing assessment. Zenodo. https://doi.org/10.5281/zenodo.345834

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In: Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.