

GESPIN 2019

11 - 13 September



This paper was presented at the 6th Gesture and Speech in Interaction Conference that was held at Paderborn University, Germany from September 11-13, 2019.

To cite this paper:

Buján, M. (2019). Gesture and speech coordination to frame utterances as humorous. In: Grimminger, A. (Ed.): *Proceedings of the 6th Gesture and Speech in Interaction – GESPIN 6* (pp. 15-20). Paderborn: Universitaetsbibliothek Paderborn. doi:10.17619/UNIPB/1-802

Gesture and speech coordination to frame utterances as humorous

Marta Buján

Universidad de Valladolid, Spain

marta.bujan.navarro@gmail.com

Abstract

The purpose of this paper is to present a multimodal study conducted on spontaneous humorous communication, in order to determine whether the pragmatic and discourse use and function of gestures and prosody differ from non-humorous communication. A sample of 14 interviews from *The Late Show with Stephen Colbert* was collected. Only interviewee's speech was analysed to ensure it was not scripted. Utterances were identified as humorous using laughter in the audience as the main criterion. The videos were annotated in ELAN for humour type, gestures (face and head movements), and prosody. The prosodic analysis was done in Praat to look into contrast between humorous and non-humorous utterances in terms of F0 and intensity. No multimodal cues specific to humour were found. The use and function of gestures in humorous utterances bear out previous studies on non-humorous communication.

1. Introduction

Humour is arguably one of the most complex instances of communication, both in terms of production and comprehension (Veale, Brône, & Feyaerts, 2015). Various studies have been conducted to look into consistent multimodal cues of humour, i.e. whether certain gestures, face expressions, head movements, changes in gaze, intonation or prosody patterns invariably associated with humour exist (Pickering et al., 2009; Attardo, Pickering, & Baker, 2011; Urios-Aparisi & Wagner, 2011; Attardo, Pickering, Lomotey, & Menjo, 2013; etc.). Many studies have been conducted on the markers of irony or sarcasm, with conflicting results (Rockwell, 2000; Attardo, Eisterhold, Hay, & Poggi, 2003; Bryant, 2010; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, and Urios-Aparisi, 2011; Tabacaru, 2014, etc.). Fewer studies exist on non-ironical humour, and fewer still focus on spontaneous humour (Archakis & Tsakona, 2005; Attardo, Pickering & Baker, 2011; Feyaerts, 2013, etc.). Nevertheless, the focus on spontaneous —non-scripted— communication is relevant, as humour is based on familiarity (Flamson, Bryant, & Barret, 2011). Given that posed humour needs to reach a wide audience, it may be delivered in an exaggerated manner, and resort to different multimodal resources from those employed in naturally-occurring, non-scripted humorous utterances (Rockwell, 2000; Urios-Aparisi & Wagner, 2011).

A survey of the literature shows that most studies have found no consistent multimodal cues of humorous speech, as compared to serious discourse. This is a counterintuitive notion, especially given the abundance of studies claiming that irony, for instance, is associated with certain intonation patterns (Rockwell, 2000; Attardo et al., 2003; Cheang & Pell, 2009; González-Fuente, Escandell-Vidal, & Prieto, 2015; etc.). These studies have yielded a wide range of often conflicting results, whereby irony is associated with flat (Haiman, 1998), rising intonation (Schaffer, 1982), higher (Rockwell, 2000) and lower pitch (Haiman, 1998; Anolli, Ciceri, & Infantino, 2000), heavy exaggerated pitch (Adachi, 1996) and relatively monotonous intonation (Haiman, 1998), etc. Attardo et al. (2003) claimed that there is no such thing as an ironic intonation, but rather that pitch and changes in prosody are just contrastive markers. Regarding gestures, Attardo, Wagner, and Urios-Aparisi (2011) compiled different ironical gestural cues appearing in the literature (Muecke, 1978; Attardo et al., 2003). Tabacaru and Lemmens (2014) argued that raised eyebrows are gestural triggers prompting the hearer to take the utterance as humorous, ironic, or sarcastic. According to González-Fuente et al. (2015), prosody and gesture are just pragmatic facilitators. For these authors, prosody and gestures, therefore, are used as tools to reduce the cognitive effort required from the hearer to interpret the ironic nature of the utterance (Yus, 2003, 2016).

This paper presents a study conducted to gain an insight into how humour is conveyed in face to face interaction. I look into how certain prosodic features, gestures, and speech interplay in the production of non-scripted humorous utterances in English to determine if functions and uses specific to humour can be found, as opposed to non-humorous communication.

2. Methods of data acquisition, annotation and analysis

2.1. Sample

The sample analysed includes 14 interviews from *The Late Show with Stephen Colbert* (Hoskin, 2015). Only utterances by interviewees have been analysed, avoiding mostly pre-scripted or rehearsed host's speech. The fully spontaneous nature of the interviewees' speech could be questioned, as most of them are people used to speaking in public and may therefore be seen as merely acting out their public persona during the show. Having said that, the aim of the research conducted for this study aimed at confronting non-scripted humorous utterances to those taken from sitcoms, TV shows or stand-up comedies in previous literature. Hence, the sample can at least be considered semi-spontaneous to the extent that it has not been previously scripted.

Each interview was analysed in a different ELAN file. Prosodic features for each selected utterance were studied separately in Praat. The sample contains 103.83 minutes of interviews, out of which 109 humorous utterances were found. For each humorous utterance, annotations on five parameters were made: a) transcription of the utterances selected, b) main construal mechanism underlying humour, c) type of humour involved, d) gestures made in the humorous utterances, e) prosodic analysis (pitch and intensity).

Following Bryant (2010), three different kind of utterances were identified with regards to the prosodic analysis: a) Humorous utterances: For the sake of objectivity, utterances were considered humorous when the audience reacted to them laughing, in order to avoid bias based on the coder's interpretation of humour and following standard practice in the literature (Morreal, 1983; Attardo, Pickering, & Baker, 2011; Archakis & Tsakona, 2005; Flamson et al., 2011; Tabacaru, 2014; Bryant & Gibbs, 2015). b) Baseline and pre-base utterances were also selected to measure prosodic contrast between humorous and non-humorous instances. Baseline utterances were those said immediately before humorous utterances, whereas pre-base were those immediately preceding baseline utterances. A control analysis could thus also be performed comparing non-humorous utterances (pre-base / baseline).

Mean pitch (F0 in Hz) and mean intensity (in dB) were obtained for each utterance. Then, all data was recorded in SPSS in order to estimate the standard deviation (SD) in mean intensity and mean pitch, for each type of utterance per interview, as a proxy measure of variability and prosodic contrast (Purandare & Litman, 2006; Bryant, 2010). SD values per type of utterance were compared within speakers through t-tests (independent variables) to determine whether there was a statistically significant difference in SD, which would lead to conclude that prosodic contrast in F0 and intensity was in turn significant. No statistically significant differences in SD values for F0 and intensity were found in the sample ($p=0.05$). Consequently, no prosodic contrast has been identified between humorous and non-humorous utterances, when it comes to F0 and intensity SD values. Admittedly, the setting and casual tone of the programme, prone to humour, would not require humour be made particularly salient through prosodic cues.

2.2. Multimodal analysis and discussion

In this section, a token of the multimodal analysis performed on the sample is included, on the basis of the most frequent combinations of humour types and gestures. The purpose of the analysis is to delve into the pragmatic and discursive use and functions of co-speech gestures in spontaneous humorous utterances to see whether differences exist with regards to non-humorous communication (Hadar et al., 1985; Poggi & Pelachaud, 1998; McClave, 2000; Kendon, 2002; Lee & Marsella, 2010; Kousidis, Malisz, Wagner, Schlangen, & Ladewig, 2013; Ishi, Ishiguro, & Hagita, 2014; Tabacaru, 2014; etc.). Only face gestures and head movements were annotated, as there was not always a clear view of the hands and the rest of the body. Data on gestures was cross-referenced with both humour types and construal mechanisms identified in the sample. No consistent

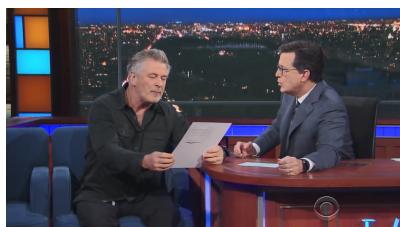
correlation patterns emerged beyond what could be expected due to the frequency of occurrence of each type of gesture, construal of humour in the sample.

Head movements and raised eyebrows have been found to serve as beats (Hadar et al., 1984; Pelachaud, Badler, & Steedman, 1996; Krahmer & Swerts, 2007; Guaiatella et al., 2009; Flecha-García, 2010; Tabacaru, 2014), that is, non-representational gestures used to punctuate speech (Kendon, 1980; McNeill, 1992). Head nods are considered to generally signal agreement (Lee & Marsella, 2010), whereas head shakes are associated with explicit or implicit negative statement (Kendon, 2002). Face gestures have been assigned various communicative functions in the literature (Poggi & Pelachaud, 1998): (1) affective display (Ekman & Friesen, 1971); (2) syntactic function, when facial expressions punctuate questions, emphasis, intonational accents, pauses, etc. (Poggi & Pelachaud, 1998); (3) dialogic function (C. Goodwin, 1980); (4) referring function (Ekman, 1979); (5) attitude display, when face gestures express the speaker's attitude towards the interlocutor (Poggi & Pelachaud, 1998).

In example (1) we find an instance of raised eyebrows and parody. Raised eyebrows have been associated with the notion of surprise, as attention-getting devices (Guaiatella et al., 2009), as tools to alert the hearer about important upcoming bits of information (Kim, Cvejic, and Davis, 2014), as underliners contributing to information structure (Flecha-García, 2010) or as gestural triggers to signal that an utterance must be interpreted as humorous. Furthermore, eyebrows have been found to strongly correlate with prosody (Flecha-García, 2010).

Stephen Colbert is interviewing Alec Baldwin, and he brings up a letter that Alec Baldwin received from President Nixon after Baldwin had lost the election for president of his school at George Washington University. Alec Baldwin then takes the letter to read what he considers to be the best part of it.

(1) Alec Baldwin: You know what the greatest part of this thing is? It's that he writes: "From our mutual friend Mark Weinberg I've learnt of the disappointing results, as far as you are concerned".



From our mutual friend...



...as far as you are concerned.

Figure 1. Raised eyebrows in example (1).

In this example, Alec Baldwin resorts to parody to delimit the part of the letter that he finds most interesting, as conveying the lack of tact by Nixon, or simply the fact that he did not feel sorry for Alec Baldwin's defeat. In order to do so, the actor mimics precisely those Nixon's words, as opposed to the first excerpt from the letter that he reads normally. The parody is shown by a change in voice quality, a significant lower pitch, a palm-up gesture, head shake, and raised eyebrows. Eyebrows are raised over the entire remark "as far as you are concerned". As I see it, in this particular instance the use of the raised eyebrows could be twofold. On the one hand, to frame the chunk of letter that Alec Baldwin considers more significant, arguably because it is the most telling part about Nixon's attitude towards his defeat, or because he feels it showed lack of empathy. On the other, the raised eyebrows could also be associated to the expression of surprise felt by Alec Baldwin on reading that part of the letter.

Most examples boast a combination of gestures co-occurring with speech. It is the coordination between modalities which ultimately serves to convey a message. For example, in Alec Baldwin's interview, upon taking his seat, right after being welcome by the host and by the audience with a very big round of applause, he thanks the audience and stresses what nice people they are. Then he utters: "It's nice and chilly in here", which elicits a bout of laughter in the audience. I posit that humour in this utterance arises from a clash in expectations about what he was supposed to say, e.g.

“it’s nice, what a warm welcome”, etc., and the fact that he actually states that it is “chilly”. Furthermore, chilly is reinforced with higher pitch, a nod and a smile.

An illustration of the importance of the interplay between gestures and speech to grasp the meaning of an utterance can be found in example (2) below, where Daniel Kaluuya is teasing Stephen Colbert by mimicking a previous remark made by the host. The humorous nature of (2) can only be understood in the context of the interview knowing what Stephen Colbert had said first, why it had been picked up by Daniel Kaluuya to mock the host, i.e. because it showcased a certain awkwardness due to racial differences, being aware that racial issues was the main topic in the film they are discussing, starred by Daniel Kaluuya. Crucially, only by seeing and listening to Daniel Kaluuya’s speech and multimodal behaviour—mimicking gestures, smile, etc.—, can the humorous intent be fully apprehended.

(2) Daniel Kaluuya: It’s like...What would I say...If I was white...What would I...?



Figure 2. Daniel Kaluuya mocking Stephen Colbert.

3. Conclusion

As briefly pointed in the examples above, the outcome of the analysis leads to conclude that the use and functions of co-speech gestures and prosody in semi-spontaneous humorous utterances in the sample is the same as in non-humorous communication.

One possible explanation of the absence of markers in humour as opposed to irony may be that prosodic cues are used only as metalanguage showing affect, that is, the position and feelings of the speaker with regards to the utterance. In the case of humour, it can be argued that there is no such detachment between the speaker and the humorous text. Both sarcastic/ironical and humorous utterances are manipulated by the speaker, but in two distinct ways. Sarcastic/ironical utterances are manipulated to show what the speaker thinks about the utterance. Humorous speech, on the other hand, is manipulated to mislead the hearer to a false interpretation to be subsequently proved wrong in order to achieve the humorous effect (Tabacaru, 2014).

Another explanation put forward to account for the difference between ironic and non-ironic humour in terms of multimodal framing associates the lack of markers to signal humour with an in-group expression of bonding on the part of the speaker, as relying on the common ground assumed to be shared with the interlocutors, and necessary for humour to be comprehended, thus demonstrating the affinity between participants (Tabacaru, 2014). Interestingly, Flamson et al. (2011) argued that as humour comprehension is influenced by context, the more background information is shared by the participants in the interaction, the less marking would be necessary for humour to be interpreted. In other words, the larger the intended audience of the humorous utterance, the more salient this humour will need to be made in order to ensure it is successfully conveyed (Attardo et al., 2003).

In light of the above, there seems to be no consistent markers of humour. Instead, prosodic and gestural cues, not specific to humour utterances, are sometimes used to communicate humour more effectively. The patterns and salience of the indices involved will eventually depend on the pragmatic context in which humour is conveyed.

References

Anolli, L., Ciceri, R., & Infantino, M.G. (2000). Irony as a game of implicitness: Acoustic profiles of ironic communication. *Journal of Psycholinguistic Research*, 29(3), 275-311.

Archakis, A., & Tsakona, V. (2005). Analyzing conversational data in GTVH terms: A new approach to the issue of identity construction via humour. *Humour: International Journal of Humour Research*, 18(1), 41-68.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, 16(2), 243–260.

Attardo, S., Pickering, L. & Baker, A. (2011). Prosodic and multimodal markers of humor in conversation in S. Attardo, M. Wagner, & E. Urios-Aparisi (Eds.), *Prosody and Humor* (pp. 224-247). Amsterdam / Philadelphia: John Benjamins Publishing.

Attardo, S., Wagner, M., & Urios-Aparisi, E. (2011). Prosody and humor. *Pragmatics & Cognition*, 19(2), 189–201.

Attardo, S., Pickering, L., Lomotey, F., & Menjo, S. (2013). Multimodality in Conversational Humor. *Review of Cognitive Linguistics*, 11(2), 402–416.

Bryant, G. (2010). Prosodic Contrast in Ironic Speech. *Discourse Processes*, 47 (7), 545-566.

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, 126(3), 1394-1405.

Ekman, P. (1979). About brows—emotional and conversational signals. In von Cranach, M. K. Foppa, W. Lepenies, & D. Ploog, (Eds.), *Human Ethology*, pp. 169-248. Cambridge: Cambridge University Press.

Ekman, P., Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), pp. 124 – 129.

ELAN (Version 5.0.0-beta) [Computer software]. (2017, April 18). Nijmegen: Max Planck Institute for Psycholinguistics.

Feyaerts, K. (2013). Tackling the complexity of spontaneous humorous interaction: An integrated classroom-modeled corpus approach. In L. Ruiz Gurillo & M. B. Alvarado Ortega (Eds.), *Pragmatics & Beyond New Series* (Vol. 231, pp. 243–268). Amsterdam: John Benjamins Publishing.

Flamson, T., Bryant, G., & Barret, H. (2011). Prosody in spontaneous humour. *Pragmatics & Cognition*, 19(2), 189-201.

Flecha-García, M. L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech Communication*, 52(6), 542–554.

González-Fuente, S., Escandell-Vidal, V., & Prieto, P. (2015). Gestural codas pave the way to the understanding of verbal irony. *Journal of Pragmatics*, 90, 26-47.

Guaitella, I., Santi, S., Lagrue, B., & Cave, C. (2009). Are Eyebrow Movements Linked to Voice Variations and Turn-taking in Dialogue? An Experimental Investigation. *Language and Speech*, 52(2/3), 207–222.

Hadar, U., Steiner, T.J., & Clifford Rose, F. (1985). Head movement during listening turns in conversation. In *Journal of Nonverbal Behavior*, 9, pp. 214 – 228.

Hadar, U., Steiner, T.J., & Grant, E.C. (1984). The timing of shifts of head postures during conversation. In *Human Movement Science*, 3, pp. 237 – 245.

Haiman, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford: Oxford University Press..

Hoskin, J. (director), & Colbert, S., Spyra, J., Stack, B., & Dinello, P. (writers). (2015-). *The Late Show with Stephen Colbert* [Television show]. Retrieved from <https://www.youtube.com/channel/UCMtFAi84ehTSYSE9XoHefig>

Ishi, C. T., Ishiguro, H., & Hagita, N. (2014). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, 57, 233–243.

Kendon, A. (2002). Some uses of the headshake. *Gesture* 2(2), 147-182.

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207-227). The Hague: Mouton and Co.

Kim, J., Cvejic, E., & Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Communication*, 57, 317–330.

Kousidis, S., Malisz, Z., Wagner, P., & Schlangen, D. (2013). Exploring annotation of head gesture forms in spontaneous human interaction. *TiGeR 2013, Tilburg Gesture Research Meeting*. Retrieved from https://pub.uni-bielefeld.de/download/2567303/2606548/tiger2013_KousidisEtAl.pdf

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396-414.

Lee, J., & Marsella, S. (2010). Predicting speaker head nods and the effects of affective information. *IEEE Transactions on Multimedia*, 12(6), 552-562.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics* 32, 855-878.

McNeill, D. (1992) *Hand and mind: what gestures reveal about thought*. Chicago & London: University of Chicago Press.

Morreal, J. (1983). *Taking Laughter Seriously*. New York: State University of New York Press.

Muecke, D. (1978). Irony markers. *Poetics*, 7(4), 363-375.

Pelachaud, C., Badler, N., & Steedman, M. (1996). Generating facial expressions for speech. *Cognitive Science*, 20, 1-46.

Pickering, L., Corduas, M., Eisterhold, J., Seifried, B., Eggleston, A., & Attardo, S. (2009). Prosodic markers of saliency in humorous narratives. *Discourse Processes*, 46(6), 517-540.

Poggi, I. & Pelachaud, C. (1998). Performative faces. *Speech Communication* (26), 5-21.

Poggi, I., D'Errico, F., & Vincze, L. (2010). Types of Nods. The Polysemy of a Social Signal. In LREC. Retrieved from http://www.academia.edu/download/4683379/596_paper.pdf

Purandare, A., & Litman, D. (2006). Humor: prosody analysis and automatic recognition for F* R* I* E* N* D* S. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 208-215. Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=1610107>

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, 29(5), 483-495.

Tabacaru, S. (2014). *Humorous implications and meanings: a multi-modal approach to sarcasm in interactional humor* (Unpublished doctoral dissertation). Université Charles de Gaulle - Lille 3. Retrieved from <http://www.theses.fr/2014LIL30015>

Tabacaru, S., & Lemmens, M. (2014). Raised eyebrows as gestural triggers in humour: The case of sarcasm and hyper-understanding. *The European Journal of Humour Research*, 2(2), 11–31.

Urios-Aparisi, E., & Wagner, M. (2011). Prosody of humor in *Sex and the City*. *Pragmatics & Cognition*, 19(3), 507-529.

Veale, T. Brône, G., Feyaerts, K. (2015). Humour as the *killer app* of language. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 1-12). Berlin/Boston: Walter de Gruyter.

Yus, F. (2003). Humor and the search for relevance. *Journal of Pragmatics*, 35(9), 1295-1331.

Yus, F. (2016). *Humour and Relevance*. Amsterdam: John Benjamins Publishing.