

Barack Obama's pauses and gestures in humorous speeches

Costanza Navarretta
University of Copenhagen
Njalsgade 136, Copenhagen
costanza@hum.ku.dk

Abstract

The main aim of this paper is to investigate speech pauses and gestures as means to engage the audience and present the humorous message in an effective way. The data consist of two speeches by the USA president Barack Obama at the 2011 and 2016 Annual White House Correspondents' Association Dinner. The success of the message is measured in terms of the immediate audience response. The analysis of the multimodally annotated data indicates that silent speech pauses structure and emphasise the discourse, and often precede the audience response. Only few filled pauses occur in these speeches and they emphasise the speech segment which they follow or precede. We also found a highly significant correlation between Obama's speech pauses and audience response. Obama produces numerous head movements, facial expressions and hand gestures and their functions are related to both discourse content and structure. Characteristics for these speeches is that Obama points to individuals in the audience and often smiles and laughs. Audience response is equally frequent in the two events, and there are no significant changes in speech rate and frequency of head movements and facial expressions in the two speeches while Obama produced significantly more hand gestures in 2016 than in 2011. An analysis of the hand gestures produced by Barack Obama in two political speeches held at the United Nations in 2011 and 2016 confirms that the president produced significantly less communicative co-speech hand gestures during his speeches in 2011 than in 2016.

1 Introduction

This paper investigates Barack Obama's use of speech pauses and co-occurring gestures as means to engage the audience and present humorous message in an effective way. Gestures are in what follows non obstructive communicative body behaviours. Speech and gestures are closely related temporally and semantically in face-to-face communication (Kendon, 2004; McNeill, 2005), and they have multiple and sometimes co-occurring functions. In particular, gestures are important signals in interaction management (Allwood et al., 1992) and they contribute to the expression of the message's content (Kendon, 2004). Similarly, speech pauses are frequent in oral communication, and they have functions which are both related to the content and the structure of the discourse (Maclay and Osgood, 1959; Goldman-Eisler, 1968; Duncan and Fiske, 1977; Shriberg, 1994; Navarretta, 2016). The speech pauses, which we include in this study are silent pauses and filled pauses, the most common being *um*, *ah*, and *uh*. The gestures we address are head movements, facial expressions and hand gestures.

The data of the study consist of two speeches by the American president Barack Obama at the Annual White House Correspondents' Association Dinner. Obama has been recognised to be a capable and elegant speaker by the press and researchers. They have especially praised the lyrical content of his discourses and the ability with which he delivers them, inter alia (Cooper, 2011). The speeches at the Annual White House Correspondents' Association Dinners are different from other presidential speeches because the president, according to the tradition, mocks himself, his collaborators, political adversaries, and the press corps.

We address speech pauses and gestures as means to engage the audience and present humorous content in an effective way and measure the success of Obama's messages in terms of the immediate audience response in the form of cheers, laughter and/or applause. We have transcribed Obama's speech, annotated his gestures and marked audience's response, and performed qualitative and quantitative analyses on these annotations.

The paper is organised as follows. In section 2, we present background literature, then in section 3 we describe the data and the annotations. Section 4 contains an analysis of the annotated data followed by a discussion in sections 5. Finally, in section 6, we conclude and present future work.

2 Gestures and Speech Pauses

The communicative functions of co-speech gestures and speech pauses are several, and both gestures and pauses are multifunctional. Co-speech gestures contribute to the content and the structure of discourse (McNeill, 1992; Kendon, 2004), and they regulate the interaction as feedback and turn management signals (Allwood et al., 1992; Sacks et al., 1974; Allwood et al., 2007). Finally, they can show the attitudinal state of the speakers and their interlocutors.

Speech pauses are voluntary or involuntary signals, which help regulating the interaction (Duncan and Fiske, 1977; Clark and Fox-Tree, 2002) and can signal that the speakers are planning and structuring their message (Maclay and Osgood, 1959; Goldman-Eisler, 1968; Shriberg, 1994; Chafe, 1987). The presence of numerous speech pauses can also indicate that the speakers are talking about difficult concepts (Reynolds and Paivio, 1968; Rochester, 1973) or are looking for the appropriate word (lexical retrieval) (Krauss et al., 2000). Hirschberg and Nakatani (1998) investigate speech pauses in read and spontaneous English speech and conclude that pauses are often used as markers of discourse structure.

Studies of the relation between speech pauses and gestures point out that they are not only temporally but also functionally related (Boomer and Dittman, 1964; Butterworth and Hadar, 1989; McNeill, 1992; Kendon, 2004; Esposito et al., 2001). In particular, Esposito and Esposito (2011) find that speech pauses often co-occur with gesture holds in English and Italian spoken data and suggest that speech pauses and gesture holds have the same function of introducing new information. More specifically, speech pauses introduce new information in the verbal modality while gestural holds introduce new information in the non-verbal one. The presence of speech pauses and gestures has also been found to contribute to the perception of naturalness of software talking agents (Cassell et al., 1994; Cassell, 2000; Maatman et al., 2005; Rehm et al., 2008).

Some studies have focused specifically on the function of speech pauses in humorous contexts. Examples are the analysis of ungrammatical silent pauses and rate of articulation in the sitcom *Friends* by Quaglio (2009) and Bilá (2014). According to them, sitcoms exhibit features of both spoken and written discourse because they are based on written texts but are acted as spontaneous speech. Speech pauses and their timing in comedy have also been addressed by both researchers and comedians. Many of these studies propose that changes in speech rate and pauses preceding punch lines are common means in humorous discourse. However, corpus-based studies by Attardo and Pickering (2011) and Attardo et al. (2011) do not confirm these assumptions. In their data, pauses do not precede punch lines and the speech rate does not change in the humorous and non-humorous parts of the conversations. Attardo et al. (2011) find though that during humorous conversations, speakers smile and laugh more often than in non-humorous conversations.

Sankey (1998) and Oliver (2013) point out that pauses in comedy are not only means to structure and emphasise the discourse, but they also give the audience time to reflect on and appreciate the conveyed message. Finally, although the importance of gestures in comedy and film is recognised, gestures in humorous discourse are often analysed independently of speech (Clayton, 2007; Weitz, 2012).

The role of speech pauses and non-verbal behaviour as communicative means in both political speeches and in humorous discourse has also been addressed by various studies. For example, Duez (1982) analyses silent and non silent pauses in French casual interviews, political interviews and televised political speeches and she reports that the total time of silent pauses is 50% longer in political speeches than in interviews, and that the longer pauses often have a stylistic function. Salvati and Pettorini (2013) analyse different Italian speeches held by Silvio Berlusconi and conclude that he uses more emphatic pauses in political speeches than in other types of discourse. Guerini et al. (2013) collect a corpus of transcriptions of American political speeches and add to the transcriptions of the speeches occurrences of audience reaction in the form of Laughter and Applause in order to find prominent discourse segments in them. In the present work, we also annotate audience reaction, but differing from Guerini et al. (Guerini et al., 2013), we also annotate Obama's speech pauses and his gestures and investigate their relation to the audience's reaction.

3 The Speeches and the Annotations

The two speeches by the U.S. president Barack Obama at the White House Correspondents' Association Annual Dinner were held in 2011 and 2016, respectively. In the rest of the paper, we refer to the speech from 2011 as *talk2011* while we refer to the the speech from 2016 as *talk2016*. The videos which we have used are the official recordings by the White House which were available at <http://www.WH.gov> while Obama was the president. In the two videos, the president is recorded frontally while he speaks as it can be seen in the two snapshots in Figure 1 and 2.

We converted the recordings from mp4 to avi format and extracted audio wav files. Silent pauses were transcribed automatically from the audio applying a PRAAT built-in script (Boersma and Weenink, 2009). We found that the best silent threshold for delimiting silent pauses in these data is -35.0 dB with the minimum silent interval



Figure 1: Snapshot from Obama's 2011 Speech



Figure 2: Snapshot from Obama's 2016 Speech

set at 0.2 seconds. Successively, Obama's speech and audience reaction were added to the TextGrid file produced by PRAAT and the automatic transcriptions of silent pauses were corrected. The resulting speech transcriptions comprise speech segments (one or more speech tokens), silent and filled pauses, and, finally, audience responses (cheers, laughter and/or applause). The PRAAT transcriptions were imported in the ANVIL tool (Kipp, 2004) and Obama's gestures were annotated in it.

We only annotated the video parts in which Obama speaks and ignored the video segments which Obama shows to the audience. The duration of the annotated *talk2011* video segments is 13 minutes and 22 seconds while the duration of the annotated *talk2016* video segments is 30 minutes.

The shape and semiotic type of the gestures were annotated following the MUMIN annotation framework (Allwood et al., 2007). Table 1 shows the shape features of head movements, facial expressions and hand gestures.

Head movements are described by the form of the movement and information about whether the movement is single or repeated. The face is described by a general face attribute and three attributes describing the shape of eyebrows, the openness of the mouth and the position of the lips. Finally, hand gestures are characterised by the following information: a) the hand(s) involved in the gesture, b) whether the gesture is single or repeated, c) the trajectory of the hand(s), d) the extension of the fingers e) the orientation of the palm at the stroke. The semiotic types of the gestures which we annotated were inspired by Peirce (1931) and comprise *indexical deictic*, *indexical non-deictic*, *iconic*, *iconic metaphorical* and *symbolic* (Allwood et al., 2007).

4 The Analysis

Table 2 contains the absolute and relative frequency (occurrence per second) of speech tokens, pauses and gestures in *talk2011*, *talk2016*, and in total. When calculating the ratio speech token per second, we excluded the time during which Obama does not speak because the audience is laughing and/or applauding. The resulting speech duration is 8 minutes for *talk2011* and 19 minutes for *talk2016*.

Table 1: Shape Features

Attribute	Value
HeadMovement	Nod, Jerk, HeadForward, HeadBackward, Shake, Waggle, HeadOther, Tilt, SideTurn
HeadRepetition	HeadSimple, HeadRepeated
General face	Smile, Laugh, Scowl, FaceOther
Eyebrows	Raise, EyebrowsOther
MouthOpen	OpenMouth, CloseMouth
MouthLips	CornersUp, CornersDown, Protruded, Retracted, LipsOther
Handedness	BothHandsSym, BothHandsAsym, RightSingleHand, LeftSingleHand
HandRepetition	Single, Repeated
Fingers	IndexExtended, ThumbExtended, AllFingersExtended, FingersOther
TrajectoryLeftHand	LeftHandForward, LeftHandBackward, LeftHandSide, LeftHandUp, LeftHandDown, LeftHandComplex, LeftHandOther
TrajectoryRightHand	RightHandForward, RightHandBackward, RightHandSide, RightHandUp, RightHandDown, RightHandComplex, RightHandOther
PalmOrientation	PalmUp, PalmDown, PalmSide, PalmVertical, PalmOther

Table 2: Absolute and Relative Frequency of Speech, Pauses and Gestures

token	talk2011 #	talk2011 #/sec	talk2016 #	talk2016 #/sec	Total	Total/sec
speech	1059	2.21	2531	2.22	3590	2.22
silent	225	0.47	243	0.21	468	0.29
filled	10	0.02	10	0.009	20	0.01
head	357	0.74	831	0.72	1188	0.73
face	66	0.14	117	0.16	183	0.15
hand	51	0.11	237	0.21	289	0.18

In order to verify whether Obama behaves differently in the two events in terms of speech rate and frequency of gestures, we tested the χ square of the observed and expected behaviour, the expected behaviour being that Obama speaks or produces a type of gesture with the same relative frequency in the two events. The difference between observed and expected behaviour is considered to be significant if the χ square p value is less than 0.001.

There is no difference in speech rate (words and fillers) in the two speeches, and this could be related to the fact that the two speeches are read as they belong to a particular genre. In *talk2011*, silent and filled pauses occur more frequently than in *talk2016*, but the difference is not statistically significant (χ square = 2.77 with 1 degrees of freedom and the 2-tailed $p = 0.096$). Silent pauses have shorter duration in *talk2011* than in *talk2016*, but also in this case the difference is not significant.

The low frequency of fillers in the two speeches was expected because Obama is reading from a manuscript. The qualitative analysis of the fillers' occurrences shows that Obama uses them consciously to give importance to what he has said or what he is going to say. In a case in *talk2016*, after having made a joke on the journalists' exaggerate coverage of Trump's campaign, Obama repeats three times the filler *hm*, and then after a longer silent pauses utters a strong *hm* in order to emphasise his point giving rise to the audience laughter.

Obama moves his head with approximately the same frequency in *talk2011* and *talk2016* (χ square equals 0.101 with 1 degrees of freedom and the 2-tailed $p = 0.751$). He moves his head continuously turning it to the right or to the left in order to address the whole audience. Sometimes, he also turns his body and talks or refers to the organisers and/or his wife who sit at the right of his podium. During speech pauses, or while the audience is laughing, Obama often moves his head forward to look at his manuscript. The latter head movements are not communicative. It must be noted that the co-occurring upper body movements were not annotated.

Also the relative frequency of facial expressions in *talk2011* and *talk2016* is approximately the same (χ square = 0.71 with 1 degrees of freedom, and the 2-tailed $p = 0.4$).

The most common facial expressions in these data are smiles, laughs and expressions in which Obama retracts his lips while listening at the audience's response or looking at his manuscript. Frowns and raised eyebrows also occur frequently in these data.

While the frequency of Obama's head movements and facial expressions does not change in the two talks, we found that he produces significantly more hand gestures in *talk2016* than in *talk2011* (χ square equals 19.295 with

1 df, and 2-tailed $p < 0.0001$). All types of co-speech hand gesture are produced in the two speeches, and the most frequent gesture types are beats and deictics. Hand gestures are also used to structure and emphasise the discourse. There are relatively more symbolic gestures in *talk2011* than in *talk2016*, and Obama produces relatively more iconic and deictic gestures in *talk2016* than in *talk2011*.

The temporal ratio of audience reaction and Obama’s speech is approximately the same in the two events. More specifically, the audience applauds and/or laughs 37% of the speech duration in *talk2011* and 40% of the speech duration in *talk2016*. The relative frequency of response per second in *talk2011* is 0.18 while in *talk2016* is 0.15. Thus, the audience laughed/applauded more often, but for shorter time in *talk2011* than they did in *talk2016*, but the difference in response frequency is not statistically significant (χ square equals 2.686 with 1 degrees of freedom, and 2-tailed $p = 0.1012$).

In some cases, Obama repeats words in order to emphasise them or prolong the audience reaction. An example of this is in *talk2011* in which he makes fun of his wife’s engagement in health food and children. It is illustrated in Table 3 in which three columns we report speech segments and pauses, co-occurring gestures and audience reaction.

Table 3: An Example from Talk2011

Obama’s speech	Co-speech gestures	Audience
<i>We made a terrific team at the Easter Egg Roll this week</i> [Pause]	Turns right and points at Michelle	
<i>I’d give out</i> [Pause]	Turns to the front	
<i>bags of candy to the kids and</i> [Pause]	Performs handing gesture with right hand (Figure 3a)	
<i>and she’d snatch them right back out of their little hands</i> [Pause]	Both hands illustrate the snatching (Figure 3b)	Audience laughter
laughs [Pause]	Turns to the right and looks at Michelle	Audience laughter
laughs	Turns to the front	Audience laughter
<i>Snatched them!</i> [Pause]	Smiles	Audience laughter



Figure 3: Snapshots of Iconic Gestures: a) Handing a Bag b) Snatching Bags

In Table 4, an example of uses of silent pauses as emphasising signals and means to get the audience react to the speech content is given. The example is from *talk2016* at a point where Obama makes fun of Hillary Clinton’s clumsiness in using the social media. Table 4 is organised as the preceding Table 3.

Finally, we measured the Pearson correlation coefficient between pause occurrences and audience response in order to confirm that pauses are related to audience reaction in these data. The correlation is positive and is highest when only silent pauses were considered (Pearson 2-tailed correlation $r = 0.465$). The correlation level is also highly significant ($r(1541) < 0.0001$).

Table 4: Example from Talk2016

Obama’s speech	Co-speech gestures	Audience
<i>You’ve got to admit though</i> [Pause] <i>Hillary trying to appeal to young voters</i> [Pause] <i>is a little bit like your relative</i> [Pause] <i>who just signed for Facebook</i> [Pause] <i>Dear America</i> [Pause] <i>did you get my poke?</i> [Pause] <i>Is it appearing on your wall?</i> [Pause]	Smiles Repeatedly moves both hands one up and one down (Figure 4)	Audience laughter Audience laughter Audience laughter

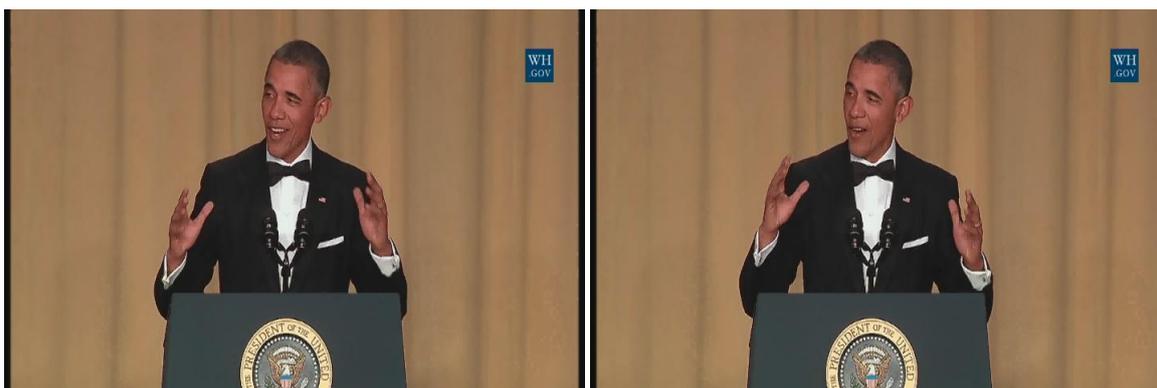


Figure 4: Snapshots of Iconic Gesture Co-occurring with Speech Segment “appearing on your wall” from *talk2016*

5 Discussion

The analysis of the multimodally annotated Obama’s speeches at the Annual White House Correspondents’ Association Dinner in 2011 and 2016 shows that Obama produces slightly more silent pauses in *talk2011* than in *talk2016*. The audience applauds more often in *talk2011* than in *talk2016* while they applaud/laugh for a longer time during the latter event. However, these differences are not statistically significant.

The analysis of silent pauses in the two speeches indicates that pauses, in most cases, delimit grammatical phrases, that is nominal, adjectival, verbal, adverbial and clausal phrases. Obama also delimits topic shifts with silent pauses, thus these data confirm what Hirschberg and Nakatani (1998) found in various types of read and spoken English data. We also found a number of pauses which preceded single words. Since Obama is mostly reading from a manuscript, these pauses cannot signal lexical retrieval as it would be the case in spontaneous speech (Krauss et al., 2000). Furthermore, they are not used to fake spontaneous speech as it is the case in sitcoms (Quaglio, 2009; Bilá, 2014). Instead, they are so-called *emphatic pauses* and emphasise the following speech segment. Finally, a number of pauses that follow a word or a phrase are used by Obama to let the audience get the point, and in numerous cases after these pauses the audience react by laughing and/or applauding the president. These emphasising use of pauses in humorous speech are described inter alia by Sankey (1998) and Oliver (2013). The correlation between pauses and audience response is positive as we expected, and therefore information about pauses should be tested as a feature for predicting audience reaction to humorous speech.

Filled pauses are few in these data, and Obama uses them consciously to emphasise what he has just said as in the case in which he repeats the same filler several times making the audience laugh. Finally, a number of short pauses follow the audience response, indicating that Obama adapts his speech to the audience’s reaction. In some cases, he controls the duration of the audience’s response by e.g. signalling with his hands that they should stop laughing/applauding or in other ways indicating that he wants to talk. In other cases, Obama makes the audience continue laughing by laughing with them or repeating parts of his preceding words as in the example illustrated in Table 3.

We did not compare Obama’s speech rate in humorous and not humorous speech and therefore cannot confirm Attardo and Pickering (2011)’s observations, but our data confirms the findings by Attardo et al. (2011) that noticed that people often laugh and smile when they deliver jokes. This was also the case for Obama who often laughs and

smiles in these speeches.

Obama often moves his head to look at his manuscript while he holds speech pauses. These gestures are not communicative, but they are typical of people who are used to present read material in a lively way. Communicative gestures only seldom co-occur with speech pauses, and this confirms the findings by Esposito and Esposito (2011) who report that gestural holds often co-occur with silent pauses and have parallel functions of introducing new gestural and verbal content, respectively.

Obama interacts continuously with his audience, talking directly and pointing to individuals in the room. Sometimes, he laughs while talking and this is often the case when he presents jokes involving his wife, Michelle, his collaborators or political adversaries. Therefore, even if many of the head movements and hand gestures produced by Obama are typical of read speeches, the many occurrences of laughter, smiles and deictics (hand gestures and head movements) pointing to people in the room, are behaviours which are not typical of political speeches. Interestingly, Obama varies the way in which he presents his jokes. Sometimes, he laughs while talking, other times he is extremely serious and therefore surprises his audience who starts laughing when they find out that he was ironic or said a joke. In these cases, Obama stops talking in order to allow the audience to understand the point. This effect of surprise in the presentation of humorous speech has been discussed *inter alia* by Beeman (1999).

Concluding, Obama uses both speech pauses and gestures to present his humorous speech and combines behaviours from normal political speeches with behaviours typical of humorous discourse. We did not find differences in speech rate and frequency of head movements and facial expressions in the two speeches, but we found significant differences in the relative frequency of hand gestures. In fact Obama moves his hands significantly more often in *talk2016* than in *talk2011*. In order to determine whether this difference in hand gesturing is a special case or reflects a development in the way Obama's presents his speeches, we annotated the communicative hand gestures in the first five minutes of two political speeches, which Obama held at the United Nations in 2011 and in 2016.¹ In the 2011 speech Obama produced 26 hand gestures (0.08 hand gestures per second) while in the 2016 he produced 68 hand gestures (0.23 hand gestures per second). The difference in occurrence frequency is statistically significant (χ square equals 18.766 with 1 df, and two tailed $p < 0.0001$) and confirms the difference in the frequency of hand gestures which we have discovered in *talk2011 talk2016*. Concluding, the two sets of Obama's speeches, which we have analysed, show clearly that Obama uses more frequently co-speech gestures in 2016 than in 2011. This indicates that Obama improves his presentation technique during his presidential term.

6 Conclusions

In this paper, we have presented a study of pauses and communicative gestures in annotated audio- and video-recordings of two humorous speeches by the American president Obama at the Annual White House Correspondents' Association Dinners in 2011 and 2016 with the aim of determining to what extent pauses and gestures contribute to the successful presentation of humorous discourse. Success of presentation was measured in terms of direct audience response. Silent pauses were automatically extracted from the audio files in PRAAT, Obama's speech and the audience reaction were transcribed and Obama's head movements, facial expressions and hand gestures were annotated.

The analysis of the data shows that Obama's speech rate and the relative frequency of head movements and facial expressions is the same in 2011 and 2016, while Obama produces significantly more hand gestures in 2016 than in 2011. An analysis of the hand gestures produced by Obama in the first part of two political speeches held at the United Nations in the same years confirms that Obama uses more frequently co-speech hand gestures in the last year of his presidency than five years earlier. Since hand gestures contribute to the presentation of the content and structure of discourse, it is evident that Obama improves his presentation technique during his two presidential terms.

Obama uses all kinds of pauses and gestures, and especially facial expressions, as means for emphasising and presenting effectively his discourse.

We found that audience responses are related to silent pauses in these data and that the correlation is significant. Therefore, we are currently testing to what extent information about silent pauses contributes to the prediction of audience response in the speeches.

Finally, it must be noted that we have not included speech content and intonation features in our analysis. These are central aspects of humorous speech, and should therefore be accounted for in the future. The type of audience, and the communicative situation are also relevant with respect to the audience reaction to jokes. These aspects should also be included in future investigations of humorous discourse.

¹The two speeches are available at <https://www.youtube.com/watch?v=UK7JEYqIfw4> and <https://www.youtube.com/watch?v=ji6pl5Vwrvk>, respectively. We analysed 306 seconds of the 2011 talk and 300 seconds of the 2016 talk excluding sequences of video frames in which the audience reaction is recorded and Obama's hand gestures are therefore not visible.

References

- Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26.
- Jens Allwood, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta, and Patrizia Paggio. 2007. The mumIn coding scheme for the annotation of feedback, turn management and sequencing. *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special Issue of the International Journal of Language Resources and Evaluation*, 41(3–4):273–287.
- Salvatore Attardo and Lucy Pickering. 2011. Timing in the performance of jokes. *Humor - International Journal of Humor Research*, 24(2):233–250.
- Salvatore Attardo, Lucy Pickering, and Amanda Baker. 2011. Prosodic and Multimodal Markers of Humor in Conversation. *Pragmatics and Cognition*, 19(2):224–247.
- William O. Beeman. 1999. Humor. *Journal of Linguistic Anthropology*, 9(1/2):103–106.
- Magdalena Bilá. 2014. A Comparative Analysis of Silent Pauses and Rate of Articulation in the Discourse of Sitcom. *Discourse and Interaction*, 7(1).
- Paul Boersma and David Weenink, 2009. *Praat: doing phonetics by computer*. Retrieved May 1, 2009, from <http://www.praat.org/>.
- D.S. Boomer and A.T. Dittman. 1964. Speech rate, filled pause and body movement in interviews. *Journal of Nervous and Mental Disease*, 139:324–327.
- B.L. Butterworth and U. Hadar. 1989. Gesture, speech, and computational stages: A reply to mcneill. In *Psychological Review*, volume 96, pages 168–174.
- Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 413–420. ACM.
- Justine Cassell. 2000. Embodied conversational interface agents. *Communications of the ACM*, 43(4):70–78.
- W.L. Chafe. 1987. Cognitive Constraint on Information Flow. In R R. Tomlin, editor, *Coherence and Grounding in Discourse*, pages 20–51. John Benjamins, Amsterdam.
- Hebert H. Clark and Jean E. Fox-Tree. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84:73–11.
- Alex Clayton. 2007. *The body in Hollywood slapstick*. McFarland.
- Marilyn M. Cooper. 2011. Rhetorical agency as emergent and enacted. *College Composition and Communication*, 62(3):420–449.
- Danielle Duez. 1982. Silent and non-silent pauses in three speech styles. *Language and Speech*, 25(1):11–28.
- S. Duncan and D.W. Fiske. 1977. *Face-to-face interaction*. Erlbaum, Hillsdale, NJ.
- Anna Esposito and Antonietta M. Esposito. 2011. On Speech and Gesture Synchrony. In Anna Esposito, Alessandro Vinciarelli, Klara Vicsi, Catherine Pelachaud, and Anton Nijholt, editors, *Communication and Enactment - The Processing Issues*, volume 6800 of LNCS, pages 252–272. Springer-Verlag.
- Anna Esposito, Karl Erik McCullough, and Frank Quek. 2001. Disfluencies in gesture: gestural correlates to filled and unfilled speech pauses. In *Proceedings of IEEE International Workshop on Cues in Communication*, Hawaii.
- Frieda Goldman-Eisler. 1968. *Psycholinguistics: Experiments in spontaneous speech*. Academic Press, London.
- Marco Guerini, Danilo Giampiccolo, Giovanni Moretti, Rachele Sprugnoli, and Carlo Strapparava. 2013. The new release of corps: A corpus of political speeches annotated with audience reactions. In Isabella Poggi, Francesca D’Errico, Laura Vincze, and Alessandro Vinciarelli, editors, *Multimodal Communication in Political Speech. Shaping Minds and Social Action: International Workshop, Political Speech 2010, Rome, Italy, November 10-12, 2010, Revised Selected Papers*, pages 86–98. Springer Berlin Heidelberg, Berlin, Heidelberg.
- J. Hirschberg and C. Nakatani. 1998. Acoustic Indicators of Topic Segmentation. In *Proceedings of ICSLP-98*, Sidney.
- Adam Kendon. 2004. *Gesture - Visible Action as Utterance*. Cambridge University Press.

- Michael Kipp. 2004. *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Ph.D. thesis, Saarland University, Saarbruecken, Germany, Boca Raton, Florida, dissertation.com.
- R.M Krauss, Y. Chen, and R. F. Gottesman. 2000. Lexical gestures and lexical access: a process model. In D. McNeill, editor, *Language and gesture*, pages 261–283. Cambridge University Press.
- RM Maatman, Jonathan Gratch, and Stacy Marsella. 2005. Natural behavior of a listening agent. In *Intelligent Virtual Agents*, pages 25–36. Springer.
- Howard Maclay and Charles E. Osgood. 1959. Hesitation phenomena in spontaneous english speech. *Word*, 15:19–44.
- D. McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, Chicago.
- David McNeill. 2005. *Gesture and thought*. University of Chicago Press.
- Costanza Navarretta. 2016. The functions of fillers, filled pauses and co-occurring gestures in danish dyadic conversations. In Linköping University Electronic Press, editor, *Postproceedings of the 3rd European Symposium on Multimodal Communication*, volume 105, pages 55–61.
- Bobbie Oliver. 2013. *The Tao of Comedy: Embrace the Pause*. Oliver.
- C. S. Peirce. 1931. *Collected Papers of Charles Sanders Peirce, 1931-1958, 8 vols.* Harvard University Press, Cambridge, MA.
- P. Quaglio. 2009. *Television Dialogue. The sitcom Friends vs. natural conversation*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Matthias Rehm, Yukiko Nakano, Elisabeth André, and Toyooki Nishida. 2008. Culture-specific first meeting encounters between virtual agents. In *Intelligent virtual agents*, pages 223–236. Springer.
- Allan Reynolds and Allan Paivio. 1968. Cognitive and emotional determinants of speech. *Canadian Journal of Psychology*, 22:164–175.
- Sherry R. Rochester. 1973. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, 2:51–81.
- H. Sacks, E. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.
- Luisa Salvati and Massimo Pettorino. 2013. A Diachronic Analysis of Face-to-Face Discussions: Berlusconi, Fifteen Years Later. In Isabella Poggi, Francesca D’Errico, Laura Vincze, and Alessandro Vinciarelli, editors, *Multimodal Communication in Political Speech. Shaping Minds and Social Action: International Workshop, Political Speech 2010, Rome, Italy, November 10-12, 2010, Revised Selected Papers*, pages 65–74. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Jay Sankey. 1998. *Zen and the Art of Stand-Up Comedy*. Routledge, New York.
- Elisabeth Shriberg. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of California, Berkeley.
- Eric Weitz. 2012. Failure as success: On clowns and laughing bodies. *Performance Research*, 17(1):79–87.