Coordination of Understanding in Face-to-Face Narrative Dialogue

Kathleen M. Eberhard (eberhard.1@nd.edu) **Hannele B. M. Nicholson** (hnichol1@nd.edu)
Department of Psychology, University of Notre Dame
Notre Dame, IN 46556 USA

Abstract

We report the results of a study investigating speakers' and addressees' coordination of understanding in face-to-face narrative dialogue. Analyses of the occurrence of addressees' acknowledgments and exemplifications of understanding showed that nonverbal forms consistently coincided with the speakers' gaze on their face. In contrast, there was less consistent correspondence between the addressees' verbal evidence of understanding and the speakers' gaze on their face. Evidence that speakers gaze off addressees' faces because of the demands of utterance planning or encoding comes from a correspondence between their gaze off the addressee's face and their production of pause fillers (*uh* and *um*), especially at the beginning of clauses.

Keywords: spoken dialogue; disfluency; gaze patterns

Introduction

Conversation is the quintessential form of language use. It is a purposeful activity requiring the coordination of two or more people. The aim of the research presented here was to examine the coordination process in a face-to-face storytelling situation by examining speakers' gaze patterns relative to delays in speaking signaled by pause fillers and relative to addressees' signals of understanding.

According to Clark (1996), language use is a joint project consisting of 4 hierarchical levels of speaker-addressee coordinated actions, which he refers to as an action ladder. Consider the case of a speaker asking an addressee, "What time is it?". At the first level, the speaker is executing a behavior that consists of producing the sounds of the utterance. The addressee, in turn, attends to the behavior (speech). At the second level, the speaker is presenting words and phrases, which the addressee identifies as such. At the third level, the speaker is signaling an intended meaning (requesting the current time), and the addressee is understanding the meaning. At the fourth level, the speaker is proposing a joint project, namely that the addressee inform him of the current time, and the addressee considers accepting the proposal. There are two essential properties of this hierarchy of actions. The first is upward causality: The actions at a lower level cause the actions at the next level up. The second property is downward evidence: Evidence of successful completion of the actions at a higher level constitutes evidence of success at all levels below it.

As Clark (1996, p. 222) states, "A fundamental principle of any intentional action is that people look for evidence that they have done what they intended to do." Furthermore, people strive to provide evidence that is sufficient for current purposes, in a timely manner, and with the least effort. In the example above, the valid, timely, and sufficient evidence of success comes from the addressee responding with the current time soon after the end of the speaker's utterance. In doing so, the addressee provides positive evidence of her acceptance of the speaker's proposed joint project at level 4 as well as positive evidence of her understanding the meaning of the speaker's utterance (level 3), her identification of the speaker's words (level 2), and her attending to the speaker's speech (level 1).

In contrast to interactive conversation, in narrative dialogues, the speaker produces sequences of utterances across an extended time, resulting in minimal turn-taking. Thus, the joint project at level 4 is an extended proposal consisting of multiple iterations through the lower 3 levels. In this situation, the highest level of evidence of successful completion is level 3 (signaling and understanding of meaning). There are two main forms of evidence that are provided by addressees. One form is *acknowledgments*, which are assertions of understanding, also referred to as backchannels (Yngve, 1970) and generic listener responses (Bavelas, Coates, & Johnson, 2000). They may be verbal, e.g., *mhm*, *okay*, *uh huh*, or non-verbal head nods. The second form of evidence is *exemplifications* of understanding, also referred to as specific listener responses (Bavelas et al, 2000). Exemplifications are reactions to the meaning of the speaker's utterance, and, as such, they constitute more valid evidence. They can be verbal, e.g., *wow*, *oh*, *that's awful*, or non-verbal, e.g., facial gestures, such as winces, grimaces, looks of surprise or sadness. Both acknowledgments and exemplifications are brief, requiring little planning and they often overlap with or occur at the end of the speaker's utterance (e.g., Goodwin, 1981).

Evidence that the addressee is attending to the speaker's execution of a communicative behavior (level 1) is provided by his or her maintaining gaze on the speaker's face (e.g., Argyle & Cook, 1976; Ehrlichman, 1981; Goodwin, 1981; Kendon, 1967). In contrast, the speaker often exhibits a pattern of gazing on and off the addressee's face (e.g., Ehrlichman, 1981, Goodwin, 1981; Kendon, 1967). In interactive conversation, this asymmetry in mutual gaze is considered one cue to turn-taking (Duncan, 1974; Kendon, 1967; Maclay & Osgood, 1959). That is, speakers typically gaze at the addressee at the end of their turn, thereby relinquishing the floor.

However, the asymmetry in speaker's and addressee's gaze on the other's face is also observed in narrative dialogues. In this situation, the speaker's gaze on and off the addressee's face is likely to reflect other aspects of

coordination than turn-taking. In particular, following Kendon (1967), Bavelas, Coates, and Johnson (2002) proposed that in narrative dialogue, speakers gaze at addressees for evidence of their understanding. Support for this proposal comes from their finding that addressees' acknowledgments and exemplifications occurred more often when speakers gazed at them than when they gazed away, and, the speakers' gaze away occurred shortly after the occurrence of this evidence of understanding.

The current study sought to replicate and extend Bavelas et al.'s (2002) findings. Specifically, like Bavelas et al., the current study tested the hypothesis that speakers gaze at their addressee's face for evidence of successful understanding (level 3), which is provided by the addressee producing verbal and/or nonverbal acknowledgments and exemplifications. The current study extends Bavelas et al's work by investigating the additional hypothesis that speakers gaze away from their addressee when the resource demands for utterance planning or encoding are high. This second hypothesis was tested by examining the co-occurrence of speakers' gaze off their addressee's face and their production of pause fillers such as *uh* or *um*, which signal a delay in speaking due to planning or encoding difficulties (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Brennan & Schober, 2001; Clark & Fox Tree, 2002; Fox Tree, 2001). Studies investigating interactive dialogues with frequent turn-taking provide some evidence for this latter hypothesis by showing that speakers often gaze away from their addressee at the beginning of their turn at talk (Kendon, 1967; Beattie, 1978), which is the point at which speakers produce pause fillers when the high demands of utterance planning and encoding cause a delay in speaking (Smith & Clark, 1993).

As mentioned above, the speaker's gaze away from the addressee at the beginning of a turn in interactive dialogues may also be a procedure by which the speaker establishes his or her turn to talk). The limited turn-taking in narrative dialogue eliminates this possible role of gaze patterns. Furthermore, we examined whether there is a co-occurrence of the speaker's gaze off an addressee's face with pause fillers that occur before a clause/discourse segment, when the demands of utterance planning should be high, as well as within a clause, where the delay may reflect lexical retrieval difficulties.

The speakers in the current study read an obscure Brothers Grimm story, which they then told to an addressee. The story involves three main characters and several subordinate characters. It also has four main scenes, corresponding to different time periods and settings. To ensure that the speakers understood the story and that they would tell it in a relatively uniform way, they completed a quiz after reading it and before telling it to the addressees. The addressees completed the same quiz after listening to the speakers tell the story to them.

Experiment

Method

Participants Seven same sex dyads participated in the experiment in exchange for payment of $10.00 each or extra credit in a course. All were native American English speaking adults with an average age of 21 years. Five dyads were female, two of which were familiar with each other prior to the experiment. One male dyad were familiar with each other prior to the experiment.

Procedure The members of the dyads signed up for an hour-long session with one person, designated as the Speaker, arriving 30 minutes before the other person, who was designated as the Addressee. Upon arriving at the lab, the Speaker was given a consent form to read and sign. Then, he or she read a printed copy of the Brothers Grimm story *Faithful John* in a quiet room. All of the participants were unfamiliar with the story prior to the experiment. After reading the story, the Speaker completed a quiz consisting of 14 multiple-choice questions about the main events and characters in the story. The questions were presented one at a time on a computer screen, and the Speaker was given as much time as needed to select a response, which was made by pressing a key on the keyboard. The Speaker was allowed to consult the printed copy of the story when answering the questions, and the correct response for each question was displayed after the Speaker made his or her response. The Speaker was then seated at a table and fitted with a free-head eye-tracker.

After the Addressee read and signed the consent form, he or she was seated at the table opposite to the Speaker. For the four dyads who were unfamiliar with each other, the Speaker and Addressee were introduced, and each was asked to tell the other about themselves (e.g., where they were living, what year they were in college, and what their major was). The dyads were then given instructions for the task. Specifically, they were told that the experiment investigates conversational interactions between two individuals, and that the Speaker was to tell a Brothers Grimm story, which she or he had just read, to the addressee. The Addressee was told that the Speaker had answered a set of comprehension questions about the story and that he or she would receive the same set of questions after listening to the Speaker tell the story to him or her. Thus, the goal was for the Addressee to understand the story sufficiently well to be able to answer the questions correctly. The Speaker and Addressee were told that they could talk to each other and that the important thing was for them to interact as naturally as possible.

The Speaker and Addressee were informed that the experimenter would remain in the room to monitor the recording equipment. However, she would have her back to them and would listen to music over headphones to prevent her from participating as an "overhearer". The Addressee was instructed to tap the experimenter on the shoulder when the speaker had finished telling the story. The entire story-

telling session was video-taped, and after it was over, the Addressee completed the quiz.

Apparatus The Speaker was fitted with an eye-tracker (Applied Science Laboratories, Model 501) consisting of a lightweight eye camera attached to an adjustable headband. The eye camera was positioned above the Speaker's left eye, and it captured an infrared image of the eye at a 60 Hz sampling rate. The distance between the centers of the corneal and pupil infrared reflections were used to calculate the relative eye-in-head position. The head band also contained a scene camera that captured an image of the Addressee's head and torso across the table. The scene camera's image was displayed on a TV monitor along with a record of the Speaker's eye movements in the form of cross hairs that were superimposed over the scene image. A brief calibration routine was conducted to map nine eye-position coordinates onto nine corresponding scene-image coordinates. The accuracy of the resulting eye fixation record was approximately 0.5° over a range of $\pm 20^{\circ}$. Lapel microphones were attached to the Speaker's and Addressee's shirts and connected to a Hi8 VCR, which also recorded the scene image and eye-movement record displayed on the TV. A Hi8 video camera, which was positioned to the side of the Addressee, recorded an image of the Speaker's head and torso. Responses on a survey administered at the end of the experimental session indicated that the eye-tracking apparatus was not distracting or only minimally distracting to the Speakers, and it was minimally to moderately distracting to the Addressees.

Video Coding The two video-taped recordings of each dyad's experimental session were digitized at a 60Hz NTSC sampling rate and aligned with each other using Final Cut Express (Apple, Inc.). The project files were annotated using frame-by-frame playback of the synchronized audio and video tracks (each frame = 33 msec). Labeled markers were inserted on the first frame of events of interest and extended to the last frame. All coding was done independently by two individuals, with a third individual (KE) reconciling any disagreements. Categories of events of interest that were marked included the following:

(1) *Speaker's gaze*: The Speakers' gaze on and off the Addressee's face was coded in a binary fashion such that the frame that marked the last consecutive fixation on the Addressee's face was followed by the frame that marked the first fixation off the face. The Speaker's gaze on the Addressee's face consisted of two or more consecutive fixations anywhere on the face. The Speaker's gaze off the Addressee's face consisted of one or more fixations in the region surrounding the face, including the Addressee's neck and torso as well as the wall behind the Addressee. In addition, the gaze off the Addressee's face included instances in which there was a loss of the eye-tracking record due to the Speaker looking down or closing his or her eyes for a period longer than a blink.

(2) *Addressee's gaze*: The first and last frames of the Addressee's gaze away from the Speaker's head were marked based on the direction of the Addressee's eye gaze available from the eye-tracker's scene image. The Addressee's gaze away typically involved looking down at the table or to the left or right of the Speaker.

(3) *Addressee's nonverbal responses*: The beginning and end frames of the Addressee's head nods (acknowledgments) and facial gestures (exemplifications) were marked. Facial gestures displayed reactions to the story's content such as surprise or disbelief in the form of eye flashes or raised eyebrows, grimaces, winces, and frowns. Smiles were not included as a nonverbal response.

Utterance Coding The Speakers' and Addressees' utterances were orthographically transcribed using Praat (Boersma & Weenink, 1996). Transcriptions of the Speaker's and Addressee's utterances were created on separate tiers in the textgrid files, with the tiers time-aligned with the digitized audio track (48 kHz sampling rate). Transcriptions were completed independently by two individuals and checked by a third (HN). The Addressee's transcriptions contained boundaries that marked the utterances' onset and offset. The utterances consisted of acknowledgments (e.g., *okay*, *mhm*, *hmmm*, *oh*, *uh huh*) and exemplifications (e.g., *wow*, *that's weird*, *crazy*), as well as requests for clarification.

Two duplicate tiers contained the transcriptions of the Speakers' utterances. One tier contained boundaries that marked intonational phrases, which typically consisted of one or two clauses. The other tier contained boundaries for individual words, which included pause fillers (e.g., *uh*, *um*) as well as silent pauses. A third tier was used to label the pause fillers with respect to whether they occurred at the beginning of a clause, within a clause, or embedded in a larger disfluency involving a repair. As shown in examples (a) and (b) below, clause-initial fillers preceded or followed one or more discourse markers (e.g., *so*, *and*, *then*, etc.). Examples (c) and (d) show fillers that occurred within a clause, and example (e) shows a filler that occurred in the middle of a larger disfluency. The numbers in square brackets show the location of a silent pause and its duration in seconds.

a.) *Clause-initial*: [0.494] um so he knows what his inheritance is except for this one [0.635]
b.) *Clause-initial*: [0.334] and [0.596] um so they know that this princess really likes gold
c.) *Within-clause*: and they take a ship across the um [0.109] sea or something
d.) *Within-clause*: and she's like wow can I [0.426] um get some of that
e.) *Mid-disfluency*: if someone sticks [0.383] um [0.227] if someone makes her lip bleed

Analyses: The markers coding the video recordings were exported from Final Cut Express and imported into Praat as labeled tiers in the textgrid files that were time-aligned with

the transcription tiers and the digitized audio track. Scripts were used to extract frequency and duration information from the tiers. The analyses of the pause fillers and gaze patterns excluded fillers that were part of a larger disfluency (i.e., the mid-disfluency fillers).

Results

As shown in Table 1, the Speakers took an average of 659 seconds, or about ten minutes, to tell the story, and they did so with an average speaking rate of 192 words per minute. The Speakers' accuracy on the quiz was slightly higher than the Addressees (average of 95% vs. 91%, respectively). For three dyads in which both the Speaker and Addressee scored less than 100%, the questions that were responded to inaccurately by the Speaker differed from the questions that were responded to inaccurately by the Addressee.

Table 1: Total time, speech rate, and quiz scores

Dyad	Total time (sec)	Words per min	S's quiz score	A's quiz score
F1	591	202	100%	100%
F2	773	175	100%	86%
F3	531	205	93%	93%
F4*	1047	178	100%	100%
F5*	627	249	93%	86%
M1	551	155	93%	100%
M2*	491	181	86%	75%
Mean	659	192	95%	91%

Note: S = Speaker, A = Addressee, F = female, M = male, * = friends prior to experiment

Table 2 shows the number and mean duration of the Speakers' gaze on and off the Addressee's face for each dyad, as well as the percentage of the total time that the Speakers' gazed off the Addressee's face. Five of the seven Speakers' exhibited the commonly reported pattern of spending more time gazing off their Addressee's face than gazing on their Addressee's face. The other two Speakers, one male (M2) and one female (F5), spent less time gazing off their Addressee's face than on it. As for the Addressees, all five female Addressees gazed at their Speaker's face 97% or more of the storytelling time. The two male Addressees gazed at their Speaker's face 79% (M1) and 46% (M2) of the storytelling time, respectively.

Table 2: Number and average duration (sec) of Speakers' gaze on and off the Addressee's face

Dyad	# Gaze on	Duration gaze on	Duration gaze off	% Total time gaze off
F1	127	1.271	3.380	73%
F2	285	1.209	1.499	55%
F3	211	1.066	1.449	58%
F4	379	0.991	1.805	65%
F5	107	5.405	0.451	8%
M1	189	0.634	2.580	78%
M2	132	2.519	1.203	32%
Mean	204	1.871	1.724	53%

Gaze and Addressees' Responses: Table 3 shows the number of the Addressees' nonverbal and verbal acknowledgments (e.g., head nods, saying *mhm*, *okay*, etc.) and exemplifications of understanding (e.g., looks of surprise, grimaces, saying *wow*, *oh my*, etc.). There was variability across the dyads in the frequency of providing evidence of understanding, with the total number of all forms ranging from 17 (M2) to 201 (F2). However, all 7 Addressees produced more acknowledgments than exemplifications as well as more nonverbal responses than verbal responses.

Table 3: Number of Addressees' acknowledgments and exemplifications and the percentage that overlapped with the Speaker's gazed on their face

	Acknowledgments		Exemplifications	
	Nonverbal	Verbal	Nonverbal	Verbal
F1	40 (75%)	9 (44%)	14 (79%)	6 (17%)
F2	111 (86%)	64 (55%)	21 (95%)	5 (60%)
F3	11 (73%)	5 (80%)	6 (83%)	0
F4	157 (60%)	28 (54%)	4 (100%)	0
F5	76 (99%)	38 (97%)	10 (100%)	11 (100%)
M1	15 (80%)	2 (0%)	0	0
M2	51 (90%)	8 (88%)	1 (100%)	5 (100%)
Mean	66 (80%)	22 (60%)	8 (93%)	4 (69%)

On average 80% of the Addressees' non-verbal acknowledgements (head nods) overlapped with the Speaker's gaze on the Addressee's face (range 60% to 99%), and 93% of their non-verbal exemplifications overlapped with the Speaker's gaze on the their face. In contrast, the average percentages of the Addressees' verbal acknowledgments and verbal exemplifications that overlapped with the Speaker's gaze on their face were less, i.e., 60% and 69%, respectively. For each dyad, the number of the Addressee's non-verbal responses and verbal responses that overlapped with the Speaker's gaze on his or her face was compared to the numbers expected to overlap by chance using the procedure described by Bavelas et al. (2002). Specifically, when the total number of nonverbal

responses or verbal responses was greater than 20, a z-value was calculated and evaluated with the normal distribution using the formula:

$$z = \frac{O - E - .5}{\sqrt{npq}}$$

where, n is the total number of responses, O is the observed number of responses overlapping with the Speaker's gaze on the face, p is the percentage of total time the Speaker spent gazing on the Addressee's face, q is 1-p, and E is the expected number of responses overlapping with a gaze on face by chance (p*n). The subtraction of .5 is a correction for continuity. When the total number of verbal or nonverbal responses was less than or equal to 20, then the combination of n, p, and O were tested for significance using the binomial distribution. The results of the tests for each dyad are given in Table 4.

Table 4: Tests of the significance of the observed number of Addressees' responses occurring with gaze on their face

Dyad	*n* total responses	*O* # with gaze on face	*p* % total time gaze on face	*z*	p-value[a]
		Addressees' Nonverbal Responses			
F1	54	41	0.27	7.95	< .0001
F2	132	116	0.45	9.81	< .0001
F3	17	13	0.42		< .002
F4	161	98	0.35	6.80	< .0001
F5	86	85	0.92	2.14	< .02
M1	15	12	0.22		< .002
M2	52	47	0.68	3.31	< .0001
		Addressees' Verbal Responses			
F1	15	5	0.27		n.s
F2	69	38	0.45	0.56	= .06
F3	5	4	0.42		< .05
F4	28	15	0.35	1.86	< .05
F5	49	48	0.92	1.27	n.s
M1	2	0	0.22		n.s
M2	13	12	0.68		< .05

[a]One-tailed test was used for binomial tests when $n \leq 20$.

Table 4 shows that the number of the Addressees' nonverbal responses that overlapped with the Speaker's gaze on their face was significantly greater than the number expected by chance for all seven dyads. In contrast, the number of the Addressees' verbal responses that coincided with the Speaker's gaze on their face was significantly greater than the number expected by chance for only three of the seven dyads, and it was marginally significant for one other dyad. The results for the nonverbal responses replicates Bavelas et al.'s (2002) findings. The current finding that the Addressees' verbal evidence of understanding less consistently overlaps with the Speaker's gaze on their face is likely due to Speaker's gaze being unnecessary for conveying this form of evidence.

Gaze and pause fillers: The Speakers produced an average of 41 pause fillers (range 15 - 76), at an average rate of 1.87 per 100 words (range 1.0 - 3.4). The correlation between the Speakers' pause filler rate and the average duration of their gaze off the Addressee's face is 0.45. As shown in Table 5, the Speakers produced more clause-initial pause fillers than within-clause ($t(6) = 4.05$, $p < .02$, two-tailed); however, clause-initial fillers were not significantly longer in duration ($t(6) = 2.08$, $p = .08$, two-tailed).

Table 5: Filled pause rate per 100 words, % of all pause fillers (number) and average duration (sec) that were clause-initial or within-clause

Dyad	Rate	Clause-initial Total	Clause-initial Dur. (sec)	Within-clause Total	Within-clause Dur. (sec)
F1	2.57	53% (27)	0.439	33% (17)	0.378
F2	3.37	67% (51)	0.399	25% (19)	0.362
F3	0.99	61% (11)	0.438	11% (2)	0.519
F4	1.45	71% (32)	0.406	20% (9)	0.340
F5	1.67	52% (22)	0.378	21% (9)	0.307
M1	2.81	55% (22)	0.422	43% (17)	0.349
M2	1.01	53% (8)	0.449	33% (5)	0.355
Mean	1.98	59% (25)	0.419	27% (11)	0.373
SD	0.94	8% (14)	0.026	11% (7)	0.068

For 6 of the 7 Speakers, all or all but one of their clause-initial pause fillers coincided with their gazing off their Addressee's face. The Speaker (F5) who spent most of the storytelling time (92%) gazing on her Addressee's face had fewer clause-initial pause fillers (23%) coinciding with a gaze off her Addressee's face than with a gaze on. Nevertheless, a binomial test of the number of the clause-initial pause fillers that coincided with her gaze off the Addressee's face was significantly greater than expected by chance ($p = .02$, two-tailed). Thus, there was a clear correspondence between the occurrence of the Speakers' gaze off their Addressee's face and their production of pause fillers at the beginning of clauses, when the demands of utterance planning and encoding re likely to be highest.

An examination of the within-clause pause fillers also provided evidence that these signals of production difficulty coincided with the Speakers' gaze off their Addressee's face in a narrative situation. Specifically, except for the Speaker (F5) who spent most of the storytelling time gazing on her Addressee's face, the number of within-clause pause fillers produced by the other six Speakers that coincided with their gaze off their Addressee's face was greater than the number expected by chance, which was calculated by multiplying the percentage of the Speaker's total time gazing off the Addressee's face by the Speaker's total number of within-clause pause fillers. Binomial tests were significant for four of the six Speakers (p-values $\leq .05$, one-tailed), and marginally significant for one Speaker (M1) ($p = .08$, one-tailed). The test was nonsignificant for the remaining Speaker (F3) due to a small number of observations i.e.,

only 2 within-clause pause fillers, both of which overlapped with the Speaker's gaze off the Addressee's face. For the Speaker (F5) who spent most of the time gazing on her Addressee's face, only 1 of her 9 within-clause pause fillers coincided with her gaze off the Addressee's face, which was equal to the number expected by chance, albeit not significant by a binomial test ($p > .05$).

Discussion

The results of the current study demonstrated coordination of understanding between Speakers and Addressees during a face-to-face narrative dialogue. Specifically, consistent with Bavelas et al.'s findings, Addressees produced nonverbal acknowledgments and exemplifications of their understanding more often when the Speaker gazed on their face than when the Speaker gazed off their face. However, across all seven dyads, there was less consistent co-occurrence of the Addressees' verbal acknowledgments and exemplifications (e.g., *mhm*, *wow*) with the Speaker's gazed on their face. This finding is likely due to the Speaker's gaze on the Addressee's face being unnecessary for conveying this evidence verbally. The results extended previous findings by providing evidence that Speakers gaze off their Addressee's face in narrative dialogues when they experience a delay in speaking due to utterance planning or encoding. Specifically, for six Speakers, nearly 100% of their pause fillers (*um*, *uh*) that occurred before a clause, when the demands of utterance planning are high, coincided with their gazing off their Addressee's face. For all seven Speakers, the number of their clause-initial pause fillers that coincided with a gaze off was significantly greater than expected by chance. There was some evidence that pause fillers that occurred within a clause also coincide with the Speaker's gaze off the Addressees' face, however this relationship was significant for only four of the six Speakers. Future research will examine the relationship between the Speaker's gaze off the Addressee's face and longer disfluent intervals, such as a syllable prolongation followed by a pause, then pause filler, etc. In addition, coordination may also be reflected in Speakers seeking and Addressees providing evidence that a disfluency involving a repair did not impede the Addressee's understanding.

Conclusion

Although there are a number of studies investigating coordination via gaze patterns, signals of understanding, and disfluencies in interactive conversation (e.g., Bard, Anderson, Chen, Nicholson, Havard, Dalzel-Job, 2007), few studies have investigated coordination in narrative dialogue. The research presented here extends previous findings by demonstrating that Speakers' gaze on and off their Addressee's face when telling a story reflect the demands of encoding meaningful messages in speech, and evidence of its success.

Acknowledgments

We thank Carlene Koken, WonJae Shin, Susan Gundersen for their assistance with data collection and coding.

References

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B. M., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory & Language, 57*(4), 616-641.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener as a collaborative process: The role of gaze. *Journal of Communication, September*, 566-580.

Beattie, G. W. (1978). Sequential temporal patterns of speech and gaze in dialogue. *Semiotica, 23*, 29-52.

Boersma, P. & Weenink, D.. (1996). Praat: A system for doing phonetics by computer. Inst. Phonetic Sci., Univ. Amsterdam, Amsterdam, The Netherlands, http://www.praat.org.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language & Speech, 44*, 123-123.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory & Language, 44*, 274-274.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73-111.

Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society, 3*, 161-180.

Ehrlichman, H. (1981). From gaze aversion to eye-movement suppression: An investigation of the cognitive interference explanation of gaze patterns during conversation. *British Journal of Social Psychology, 20*, 233-241.

Fox Tree, J. E. (2001). Listeners' uses of um & uh in speech comprehension. *Memory & Cognition, 29*, 320-326.

Goodwin, C. (1981). *Conversational Organization: Interactions between Speakers and Hearers*. New York: Academic Press.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*, 22-63.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19-44.

Smith, V. L. & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory & Language, 32*, 25-38.

Yngve, V. H. (1970). On getting a word in edgewise. *Papers from the sixth regional meeting of the Chicago Linguistics Society* (pp. 567-578). Chicago: Chicago Linguistic Society.