Who's next? Speaker-selection mechanisms in multiparty dialogue

Volha Petukhova
Tilburg Center for Creative Computing
Tilburg, the Netherlands
v.petukhova@uvt.nl

Harry Bunt
Tilburg Center for Creative Computing
Tilburg, the Netherlands
harry.bunt@uvt.nl

Abstract

Participants in conversations have a wide range of verbal and nonverbal expressions at their disposal to signal their intention to occupy the speaker role. This paper addresses two main questions: (1) How do dialogue participants signal their intention to have the next turn, and (2) What aspects of a participant's behaviour are perceived as signals to determine who should be the next speaker? Our observations show that verbal signals, gaze redirection, lips movements, and posture shifts can be reliably used to signal turn behaviour. Other cues, e.g. head movements, should be used in combination with other signs in order to be successfully interpreted as turn-obtaining acts.

1 Introduction

Turn management is an essential aspect of any interactive conversation and involves highly complex mechanisms and phenomena. Allwood (2000) defines turn management as the distribution of the right to occupy the sender role. People do not start or stop talking just anywhere, and not without a reason. The decision to take the next turn or to offer the next turn to the partner(s) depends on the speaker's needs, motivations and beliefs, and on the rights and obligations in a conversational situation.

In the widely quoted study of Sacks, Schegloff and Jefferson (Sacks et al., 1974) a model for the organisation of turn-taking in informal conversations has been proposed. The authors observed that conversations most often proceed fluently, that mostly one conversational partner talked at a time, that occurrences of more than one speaker at a time were brief, and that transitions from one turn to the next without a gap or overlap were very common. They reasoned that there must be an underlying system of turn-taking involved in conversations. They posited that during a conversation there are natural moments to end a turn and initiate a new one, called Transition Relevance Places (TRPs), and formulated the following rules:

- If the current speaker (S) selects the next speaker (N) in the current turn, S is expected to stop speaking, and N to speak next.
- If S's behaviour does not select the next speaker, then any other participant may self-select. Whoever speaks first gets the floor.
- If no speaker self-selects, S may continue.

The generality of these rules makes them explanatory and applicable in many situations, but prevents them from being specific about the characteristics of speaker-selection techniques. At least two questions remain: (1) Which perceived behavioural aspects are used by people to estimate the locations of TRPs, and (2) Which aspects of communicative behaviour serve as signals to determine who is a potential or intended speaker of the next turn.

With respect to the first question, recent years have seen a number of solid qualitative and quantitative findings. It was observed that many turn transitions happen without temporal delays because a potential next speaker knows when a turn ends. People are able to predict turn endings with high accuracy using semantic, syntactic, pragmatic, prosodic and visual features (Ford & Thompson, 1996; Grosjean & Hirt,1996; De Ruiter et al., 2006; Barkhuysen et al., 2008, among others).

While end-of-turn prediction has been studied extensively, little research has been done on the prediction who is a potential next speaker, and on next speaker self-selection behaviour. This is in particular important if we deal with more than two participants in dialogue. Dialogue participants may just start speaking if they want to say

something, but they often signal their willingness or readiness to say something. In other words, they perform certain actions to take the turn over. Speakers may signal that they want to have the turn when it is available (*turn taking*); that they want and are ready to have the turn when it is given to them by the previous speaker (*turn accepting*); and that they want to have the turn despite the fact it is not available (*turn grabbing*).

In this study we focus on the properties of a speaker's utterances that correlate with his turn-obtaining efforts in multi-party dialogue. Correlation indicates that two variables are related, but does not measure cause. It does not mean that signs which are correlated with turn-obtaining efforts are interpreted as such by communicative partners. To investigate this issue, we also looked if speaker changes really occur shortly after certain signals have been sent. We should also take into account, however, that a participant's wish to have the turn may be overlooked or ignored by others for some reason, and that he does not get the opportunity to speak. Therefore, to obtain more certainty about utterance properties related to turn taking, we performed perception experiments where subjects judged the participant's turn-taking efforts.

Before discussing our analysis and findings we first introduce a few concepts and terms for the rest of this paper. The term 'turn' is used in the literature in two senses: (1) as in 'to have the turn', i.e., to occupy the speaker role; and (2) to refer to a stretch of communicative behaviour produced by one speaker, bounded by periods of inactivity of that speaker or by activity of another speaker. Turns in this sense are sometimes called 'utterances' (cf. Allwood, 2000). We will use the term 'turn' in this paper in both senses, in such a way that no confusion is likely to arise. A turn in the latter sense may contain several smaller meaningful parts, most often called 'utterances'; these units are linguistically defined stretches of communicative behaviour. In natural spoken dialogue, the stretches of communicative behaviour that have a communicative function do not always coincide with turns or utterances, since they may be discontinuous due to the occurrence of filled and unfilled pauses, self-corrections, restarts, and so on; and they may spread over multiple turns, when the speaker provides complex information which he divides into parts in order not to overload the addressee. The notion of *functional segment* was therefore introduced, defined as the smallest (possibly discontinuous) stretch of communicative behaviour that has a communicative function (and possibly more than one) (Geertzen et al., 2007). The notion of functional segment is especially useful when analysing the turn-taking behaviour of participants in dialogue because it allows multiple functional segments that are associated with a specific utterance or turn to be identified more accurately.

The rest of this paper is organized as follows. After introducing the corpus and its annotation in Section 2, we discuss our observations concerning the turn-taking behaviour of dialogue participants. Section 3 describes perception experiments, and reports on the recognition of a participant's behaviour as a turn-management signal. Conclusions are drawn in Section 4.

2 Observation study

2.1 Corpus material and annotations

In this study we used human-human multi-party interactions in English (AMI-meetings).[1] The *AMI corpus* contains manually produced orthographic transcriptions for each individual speaker, including word-level timings. Two scenario-based[2] meetings were selected with a total duration of 51 minutes, constituting a corpus of 2,396 functional segments which contain either verbal components, nonverbal components, or both. All four participants were English native speakers.

The nonverbal behaviour of the dialogue participants was transcribed using video recordings for each individual participant, running them without sound to eliminate the influence of what was said. This transcription includes gaze direction; head movements; hand and arm gestures; eyebrow, eyes and lips movements; and posture shifts. Transcribers were asked to annotate low-level features such as form of movement (e.g. head: nod, shake, jerk); hands: pointing, shoulder-shrug, etc.[3]; eyes:

[1] Augmented Multi-party Interaction (`http://www.amiproject.org/`).

[2] Meeting participants play different roles in a fictitious design team that takes a new project from kick-off to completion over the course of a day.

[3] Hand gesture transcription was performed according to Gut,U., Looks, K., Thies, A., and Gibbon, D. (2003). CoGesT: Conversational Gesture Transcription System. Version 1.0. Technical report. Bielefeld University `http://www.spectrum.uni-bielefeld.de/modelex/publication/techdoc/cogest/`

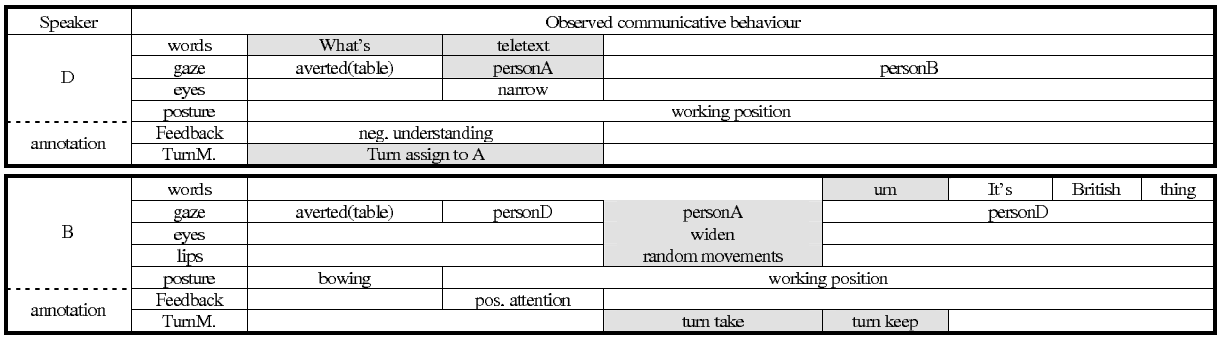

Figure 1: *Transcription and annotation example.*

narrow, widen; lips: pout, compress, purse, flatten, (half)open, random moves); direction (up, down, left, right, backward, forward); trajectory (e.g. line, circle, arch); size (e.g. large, small, medium, extra large); speed (slow, medium, fast); and repetitions (up to 20 times). The floor transfer offset (FTO: the difference between the time that a turn starts and the moment the previous turn ends) and duration of a movement (in milliseconds) were computed. At this stage no meaning was assigned to movements.

For each token in verbal segments prosodic features were computed. Prosodic features that are included are pause before the token, minimum, maximum, mean, and standard deviation of pitch (F0 in Hz), energy (RMS), voicing (fraction of locally unvoiced frames and number of voice breaks), speaking rate (number of syllables per second) and duration of the token. We examined both raw and normalized versions of these features[4]. For each verbal segment FTO, duration and word occurrence[5] features were computed.

Speech and nonverbal signs were annotated with the DIT^{++} tagset[6] using the ANVIL tool[7]. Utterances were segmented per dimension according to the approach presented in (Geertzen et al., 2007). For turn management DIT^{++} distinguishes between turn-obtaining acts (turn-initial acts) and acts for keeping the turn or giving it away (utterance-final acts). A turn-initial function indicates whether the speaker of this turn obtains the speaker role by grabbing it (*turn grab*), by taking it when it is available, (*turn take*) or by accepting the addressee's assignment of the speaker role to him (*turn accept*). A turn ends either because the current speaker assigns the speaker role to the addressee (*turn assign*), or because he offers the speaker role without putting any pressure on the addressee to take the turn (*turn release*). A turn may also have smaller units with boundaries where a reallocation of the speaker role might have occurred, but does not occur because the speaker indicates that he wants to keep the turn. Such a segment has a *turn keep* function. A segment was labelled as having a turn-management function only if the speaker performed actions for the purpose of managing the allocation of the speaker role. For example, a segment was annotated as having the function Turn Take only if the speaker performs a separate act to that effect. If the speaker just goes ahead and makes a contribution to the dialogue, without first signalling his intention to do so, then the segment was not marked with a Turn Management function. 412 segments were identified having a turn-initial function (17.2%) and 370 segments as having one of the turn final functions (15.4%). Figure 1 provides an example from the annotated corpus.

We examined agreement between annotators in identifying and labelling turn management segments using Cohen's kappa measure (Cohen, 1960)[8]. Two annotators who were experienced in

[4]Speaker-normalized features were obtained by computing z-scores (z = (X-mean)/standard deviation) for the feature, where mean and standard deviation were calculated from all functional segments produced by the same speaker in the dialogues. We also used normalizations by the first speaker turn and by prior speaker turn.

[5]Word occurrence is represented by a bag-of-words vector (1,640 entries) indicating the presence or absence of words in the segment.

[6]For more information about the tagset, please visit: `http://dit.uvt.nl/`

[7]For more information about the tool visit: `http://www.dfki.de/~kipp/anvil`

[8]This measure of agreement takes expected agreement into account and is often interpreted as follows: 0=none; 0-0.2=small; 0.2-0.4=fair; 0.4-0.6=moderate; 0.6-0.8=substantial; and 0.8-1.0=almost perfect.

annotating dialogue and were thoroughly familiar with the tagset reached substantial agreement (kappa = .76) in identifying turn segments and assigning turn-management functions.

2.2 Results

It was observed from the annotated data that meeting participants often indicate explicitly when they wish to occupy a sender role. More than half of all speaker turns were preceded by attempts to gain the turn, either verbally or nonverbally (59%). 17.2% of all functional segments were found to have one of the turn-initial functions: 12% are turn-taking segments, 4.4% have a turn-grabbing function and 0.8% are turn accepts. Consider the following examples:

(1) *B: What **you guys** received?* (Turn Release)
*A1: 0.54 **Um**(0.65)* (Turn Take) [9]
A2: I just got the project announcement

(2) *B1: yeah brightness and contrast*
*D1: - 0.35 **Well**0.19* (Turn Grab)
D2: 0.11 what we're doing is we're characterizing

(3) *B1: That something we'd want to include*
*B2: do **you**(participant D is gazed) think?* (Turn Assign)
*D1: 1.82 **Uh**(1.39)* (Turn Accept)
D2: Sure

The reasons to take the turn may be various. First, a participant may have reasons to believe that he was selected for the next turn by the previous speaker. This puts a certain pressure on him to either accept the turn or signal its refusal. Second, a dialogue participant may want to make a contribution to the dialogue and believe that the turn is available. Finally, a dialogue participant may wish to have the turn while believing that it is not available, because (1) he has a desire to express his opinion urgently; or (2) he wants to gain control over the situation, e.g. when the meeting chairman needs to get a grip on the interactive process; or (3) he notices that the current speaker is experiencing difficulties in expressing himself, and e.g. assists in completing the utterance; or (4) he wants to express his appreciation of an idea or suggestion put forward by another participant; or (5) he failed to process the previous utterance of another participant and needs immediate clarification; or (6) he expects the current speaker to finish his utterance, and wishes to be the next speaker before the partner completes his turn.

Verbally, turn-taking intentions were mainly expressed by the following tokens: *um* and its combinations such as *um okay*, *um alright*, *um well* and *um yeah* (11.5% of all turn-initial segments); *so* (5%); *and* and combinations like *and so, well and,* also by *um and*, *uh and*, *and um*, *and uh* (7.9%); *well* (5.8%); *right* and combinations like *right so* and *right well* (7%); *uh* (5.6%); *okay* and *mm-hmm*/*uh-uhu* (5%); *alright* (2.8%); *yeah* or its repetition (15.7%); *but* (2%); *just* (1.2%); and repetitive expressions (e.g. *I.. I.. I.. would like*) (1.5%).

The majority of these tokens may serve several communicative functions is dialogue. For example, *'um'* and *'uh'* are known to be used as fillers to stall for time and keep a turn. Moreover, these tokens also occur in segments which are not related to turn management. For example, *'okay'* can be used as positive feedback or to express agreement. They also can be multifunctional expressing, for example, positive feedback and turn taking simultaneously. Previous studies, e.g. (Hockey, 1993) and (Gravano et al., 2007), confirmed that the use of these cue phrases can be disambiguated in terms of position in the intonation phrase and analysis of pitch contour.

We observed significant mean differences between turn-initial use and non-turn-initial use of these tokens in terms of duration (turn-initial tokens being more than 115 ms longer); mean pitch (turn takings have $>$ 12Hz); standard deviation in pitch ($>$ 5Hz); and voicing (5% more voiced). As for temporal properties of verbal turn-initial functional segments, it was observed that the floor transfer offset (FTO) is between -699 and 1030 ms, where negative value means overlap and positive a gap between successive turns. Turn-grabbing acts have an FTO from -699 to -166ms; turn-accepting acts may also slightly overlap the previous segment and have FTO from -80ms to 136ms; turn-taking acts the longest FTO have (between 582 to 1030ms).

To assess the importance of nonverbal signs for identifying turn-initial segments, we conducted a series of correlation tests using the phi-coefficient. The phi measure is used to test the relatedness of categorical variables, and is similar to the correlation coefficient in its interpretation. Table 1 shows the correlation between segments annotated

[9] Here and elsewhere in the text figures given between brackets in examples indicate token duration in seconds; figures without brackets indicate silences between tokens in seconds.

(Non-)verbal signal	ϕ
wording (presence of tokens listed above)	.47*
any gaze redirection	.79*
direct-averted	.42*
direct(>1 person)-averted	.61*
head movement	.05
hand/arm movement	.01
eye shape change + eyebrow movement	.15
any lips movement	.59*
half-open mouth	.39*
random lips movements	.28*
posture shift	.87*
working position-leaning backward/forward	.29*

Table 1: Nonverbal signals correlated to turn-initial segmets (* significant according to two-sided t-test, $< .05$)

as having a turn-initial function and accompanying nonverbal signals.

Strong positive correlations were observed for gaze aversion, lip movements and posture shifts. Especially in multi-party conversations gaze plays a significant role in managing fluent turn transitions than in two-person dialogues, because of the increased uncertainty about who will be the next speaker. As for gaze patterns that accompany turn-initial segments, in 29.4% of the cases the participant has direct eye contact with his addressee. In 11.8% of the cases the participants who want to have the next turn gazes at more than one of the partners, most probably verifying their intention concerning the next turn. A dialogue participant who aims for the next turn first gazes at one or more partners, and averts his gaze shortly before starting to speak (44.1%).

Comparable patterns were observed in previous studies. A speaker usually breaks mutual gaze while speaking and returns gaze to the addressee upon turn completion (Kendon, 1967). Goodwin in (1981) claims that the speaker looks away at the beginning of turns and looks towards the listeners at the end of the turn. More recently, Novick (1996) found that 42% of the turn exchanges follows a pattern in which the speaker looks toward the listener while completing the turn. After a short moment of mutual gaze the listener averts his gaze and begins the next turn.

Independent from the possible meanings of specific types of head movements, and from their feedback functions, head movements are used for turn management purposes. It was noticed in (Hadar et al., 1984) that speakers use head movements to mark syntactic boundaries and to regulate the turn-taking process. In our data the intention to have the next turn was successfully signalled by repetitive short head movements (34.3%). In 11.8% of the cases turn-initial efforts were signalled by waggles (head movement back and forth and left to right) and often indicated negative feedback or uncertainty. In 3.9% of the cases headshakes as signals of disagreement were observed. Interestingly, however, head movements do not correlate significantly with turn-initial acts. By contrast, a combination of spoken signals like 'okay' or repetition of 'yeah' and multiple head nods are good signals of a participant's turn-obtaining intention (ϕ=.41, p =.003). This is in accordance with Jefferson's findings that people proceed from 'mm-hmm' to 'yeah' when they want to have the turn (Jefferson, 1985).

Hand and arm gestures that may be related to the participant's intention to have the turn were not observed frequently. We identified some shoulder shrugs that signalled uncertainty (3.5%) accompanied by head waggles and hand movements when a participant listening to the speaking partner suddenly moves his hand/fist away from the mouth (2%) or makes an abrupt hand gesture for acquiring attention (3.9%).

To signal the intention to have the next turn, participants frequently made random silent lip movements, compressing, biting, licking, or pouting their lips (10.9%). They also often keep their mouth (half-) open (47.3%). In 16.4% they narrow (possible sign of negative feedback) or widen (indicating surprise) their eyes accompanied by lowering or raising eyebrows, respectively.

Various types of upper-body posture shifts were often used as turn-initial signals (25.5%). Participants would change their body orientation from working position (both hands on the table, leaning slightly forward, head turned to the speaker) to leaning forward, backward or aside (17.6%), producing random shifts (shifting one's weight in a chair) in 2%, shifting from bowing position (bending, curling, or curving the upper body, usually while writing) (5.9%). Cassell et al. in (2001) looked at posture shifts at turn boundaries and discourse segment boundaries, and showed that both boundaries had an influence on posture shifts. Posture shifts with the upper body were found more frequently at the start of a turn than in the middle or end (48%, 36%, and 18% respectively).

Generally, dialogue participants recognize an intention to take the turn successfully. In 60.8% of all the cases turn-obtaining efforts were acknowledged and the partner's wish to have the turn was satisfied. Participants who used more than one turn-initial signal or two modalities (e.g. combining head movements and posture shifts, or verbal and nonverbal signs) were more successful in obtaining the next turn. As for the remaining 39.2% it is difficult to judge whether the turn-taking efforts were interpreted as such by partners and ignored, or whether the signals were overlooked. Looking closer at gaze behaviour of meeting participants, our intuition is that in the majority of cases (65.2%) the turn-gaining efforts were most probably overlooked, because the participant was not gazed at by other partners. In another 34.8% of the cases, the participant's turn-gaining efforts were most likely ignored, since the partners did have direct eye contact. Nonetheless, since our analysis is based on the interpretation of annotators, this intuition could be wrong. To deal with this problem, perception experiments were performed which are reported in the next section.

3 Perception study

3.1 Stimuli and procedure

Two series of perception experiments were designed to study whether naive subjects interpreted certain behaviour of meeting participants as signals to have the next turn. From the annotated data we randomly selected 167 video clips with 4 different speakers (2 male, 2 female). Two referees judged the clips assigning them to the following categories:

1. a turn-initiating act is performed when the next turn is available;
2. a turn-initiating act is performed when the next turn was assigned to this participant;
3. a turn-initiating act is performed when the turn is not available but the participant needs: (a) to signal negative feedback on processing the partner's utterance; (b) to elaborate the partner's utterance; (c) to address the partner's suggestion; (d) to clarify the partner's utterance; or (e) to shift the topic;
4. no turn-taking act is performed.

The judges reached a substantial agreement on this task (kappa scores of .67). 52 stimuli, on which the judges fully agreed, were selected for further experiments: 4 of category 1; 4 of category 2; 36 of category 3; and 8 of category 4. The duration of each clip was about 10 seconds, containing the full turn of the previous speaker, and the recordings of the participant's movements and pause after the turn (if any) till the next turn starts. The subjects had 10 seconds to react to each stimulus. They were given the task to answer the question whether they think that a participant in question is performing any turn-initial act or not.

	without sound	**with sound**
turn take	.31	.65
turn accept	.20	.55
turn grab	.32	.43
no turn-initial act	.79	1.00
overall	.48	.64

Table 2: Cohen's kappa scores for each class label for two sets of rating experiments

15 subjects (4 male and 11 female, all between the ages of 20 and 40) participated in one of the two sets of experiments: 9 subjects were asked to evaluate the video fragments without sound and 6 subjects evaluated the same fragments which were provided with sound. They were allowed to watch every video as many times as they liked.

3.2 Results

3.2.1 Subject rating

We examined inter-subject agreement using Cohen's kappa measure (Cohen, 1960). Table 2 shows kappa scores calculated for each individual condition, for two class labels and for two sets of ratings.

Subjects reached moderate agreement judging whether a meeting participant performed a turn-initial act or not if they could not hear what was said, relying only on their interpretation of the nonverbal information; they reached substantial agreement when they could hear what was said. Agreement is higher (.79 = substantial agreement when judging videos without sound, and 1.00 = perfect agreement when sound was available) when a participant does *not* display any turn-taking efforts. Among the turn-initial acts the turn grabbing which was performed to signal negative feedback on the previous speaker utterance (at the level of interpretation or of evaluation) has been evaluated with higher agreement than the others (.57, $t < .05$) under both condition, most probably because participants produce distinctive facial expressions characterized by changing an eye shape,

eyebrow and lips movements, often accompanied by a head shake or waggle additionally to other signals. The lowest agreement was found rating the turn-accept efforts of dialogue participants. This can be explained by the fact that participants to whom the next turn is assigned do not necessary perform any extra (nonverbal) action to indicate that they wish to be the next speaker, so that the raters often judge the participant's behaviour as having no turn-management function if they cannot hear that the turn was actually assigned by the previous speaker. Raters who could hear what the other participants say reached higher agreement than judges to whom speech transcription was not available. Thus, context information, such as the previous speaker's turn, seems to be important for the perception of turn-taking behaviour, perhaps also because dialogue participants actually anticipate TRPs (Ruiter et al., 2006), which makes it easier to perceive speaker-selection actions and to interpret turn-obtaining intentions.

3.2.2 Recognition of turn-initial acts

In this section we describe nonverbal features which we think may be helpful for explaining why subjects interpreted a participant's behaviour as having a turn-obtaining function (or not). We examined the following features: (1) gaze (directed, averted and combination of those); (2) head movement, if any; (3) hand gesture, if any; (4) eyebrow movement, if any; (5) eye shape change, if any; (6) lips movement, if any; (7) posture shift, if any; and (8) some combinations of these features.

We conducted a series of statistical tests, similar to those described in Section 2.2, and measured for each class label the correlations between the proportion of subjects that chose each label and the features described above. Table 3 presents correlations for the conditions with and without sound.

We can conclude that nonverbal signals are important for recognizing speaker-selection intentions. A gaze pattern such as 'gazing at more than one person and then averting the gaze', and various types of lips movements and (half-)open mouth in particular, correlate positively with a turn-initial act and have strong negative correlation with non-turn-initial acts). Head nods, on the other hand, turn out not to be significant for turn-taking purposes, because they may be used to signal active listening without the intention to take the turn (e.g. so-called backchannels). A combination of head movements and other signals, by contrast, was perceived by judges as a turn-initial signal, e.g. a head movement accompanied by lips movements, or posture shifts and certain gaze pattern such as 'mutual gaze - averted' (the combination of all three has a strong positive correlation with turn-initial acts: .55, $t < .05$). Thus, dialogue participants who use multiple signals or modalities are more successful in gaining the next turn. Conversational partners are also more likely to perceive and understand the partner's turn behaviour when relying on multiple information sources.

	ϕ **(without sound)**	ϕ **(with sound)**
turn-initial act		
gaze 'averted'	-.34*	-.44*
gaze 'direct(more persons)-averted'	.54*	.52*
head movement	.49	.25
head nods	.40	.28
hand gesture	.49	.21
eye shape change + eyebrow movements	.54*	.46*
(half-) mouth	.58*	.35*
lips movement	.44	.34*
posture shift	.41	.30*
'posture shift + head movement'	.34	.35*
'lips + head movements'	.57*	.39*
'eye shape change + head movements'	.47	.27
'eyebrow + head movements'	.46	.25
'gesture + head movements'	.44	.15
gaze 'direct-averted' + posture shift	.37	.34*
gaze 'direct-averted' + head movement	.55*	.40*
gaze 'direct-averted' + lips movements	.60*	.59*

Table 3: Features correlated with the proportion of votes for each class label (without/with sound ratings). * differs significantly from zero according to two-sided t-test, $t < .05$

4 Conclusions and future work

In this study we were interested in identifying speaker-selection mechanisms in multiparty dialogue. The main aim was to determine which aspects of a participant's behaviour serve to signal the intention to have the next turn.

A range of verbal expressions may be used to signal the intention to have the next turn, including several types of fillers, discourse markers, repetitive expressions, and other vocal sounds.

We have found that gaze redirection is the most important nonverbal indicator of turn management in multiparty dialogue, although turn organisation cannot be explained completely by gaze behaviour. In general, a participant who wants to claim the next turn first looks at the other participants and averts his gaze shortly before starting to

speak.

As for head movements, multiple head nods were found to be significantly correlated with turn-initial acts. The results of the perceptual study showed, however, that head nods are not interpreted as having a turn-initial function. By contrast, some combinations of head movements and other signals, either verbal ('okay' or 'yeah') or nonverbal (e.g. lips movements) are associated with turn-initial functional segments.

Concerning hand and arm gestures, no statistically significant results can be reported due to the low frequency of their occurrence in our data.

According to our data, facial expressions are used not only to express emotions, attitudes and states of cognitive processing, but also the intention to occupy the speaker role. Our observational and perceptual analyses show that lips movements and changes in eye shape correlate positively with turn-initial acts.

Posture shifts, finally, were frequently found at the start of a turn, and strongly correlate with turn-initial acts; they were perceived as a strong turn-initial cue on their own and in combination with other signals.

From our observational and perceptual studies it may be concluded that the combination of nonverbal signs and signals from several modalities (speech and movements) forms a reliable indicator of the intention to take the turn, and the dialogue participants who used these complex signals for the purpose to claim the next turn were successful in getting it.

This paper reports results from a limited number of dialogues and small-scale perceptual experiments, but the findings are promising. Future research will look into the performance of perceptual experiments with richer sets of stimuli, and use the results also for further observational analysis, since it is still very hard to obtain high-quality annotated data of nonverbal behaviour.

References

Jens Allwood. 2000. An activity-based approach to pragmatics. In: Harry Bunt and William Black, editors, *Abduction, Belief and Context in Dialogue; Studies in Computational Pragmatics*, John Benjamins, Amsterdam, The Netherlands, pp. 47–80.

Pashiera N. Barkhuysen, Emiel J. Krahmer, and Mark G.J. Swerts. 2008. The interplay between auditory and visual cues for end-of-utterance detection. *The Journal of the Acoustical Society of America*, 123(1):354–365.

Harry Bunt. 2006. Dimensions in dialogue annotation. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*.

Justine Cassell, Yukiko I. Nakano, Timothy W Bickmore, Candace L. Sidner, and Charles Rich. 2001. Non-Verbal Cues for Discourse Structure. In: *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, Toulouse, France.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20: 37–46.

Cecilia E. Ford and Sandra A. Thompson. 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: Emanuel A. Schegloff and Sandra A. Thompson, editors, *Interaction and grammar*, Cambridge: Cambridge University Press, pp. 135– 184.

Jeroen Geertzen, Volha Petukhova and Harry Bunt. 2007. A Multidimensional Approach to Utterance Segmentation and Dialogue Act Classification. In: *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, Antwerp, pp. 140–149.

Francois Grosjean and Cendrine Hirt. 1996. Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects. *Language and Cognitive Processes*, 11:107–134.

Charles Goodwin. 1981. Conversational Organization: Interaction between hearers and speakers. New York: Academic Press.

Agustin Gravano, Sefan Benus, Hector Chavez, Julia Hirschberg and Lauren Wilcox. 2007. On the role of context and prosody in the interpretation of 'okay'. In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic.

Uri Hadar, Timothy J.Steiner, Ewan C. Grant, and F. Clifford Rose. 1984. The timing of shifts of head postures during conversations. *Human Movement Science*, 3:237–245.

Beth Ann Hockey. 1993. Prosody and the role of okay and uh-huh in discourse. In: *Proceedings of the Eastern States Conference on Linguistics*, pp. 128-136.

Gail Jefferson. 1985. Notes on a systematic Deployment of the Acknowledgement tokens 'Yeah' and 'Mmhm'. *Papers in Linguistics*, 17(2): 197–216.

Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologia*, 26: 22–63.

David G. Novick, Brian Hansen, and Karen Ward. 1996. Coordinating Turn-taking with Gaze. In: *Proceedings of the International Symposium on Spoken Dialogue*, Philadelphia, PA, pp.53–56.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4): 696–735.

Jan Peter de Ruiter, Holger Mitterer, and Nick J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82: 515–535.