

Prosodic turn-yielding Cues with and without optical Feedback

Caroline Clemens

Technische Universität Berlin
Deutsche Telekom Laboratories
Ernst-Reuter-Platz 7
10587 Berlin

Caroline.Clemens@telekom.de

Christoph Diekhaus

Technische Universität Berlin
Department f. Language and Communication
Straße des 17. Juni 135
10623 Berlin

c.diekhaus@alice-dsl.net

Abstract

The authors present a study of prosodic turn-taking indicators. The aim was to investigate whether some of the prosodic cues increase in quality or quantity if the optical feedback channel in the verbal conversation is missing. For the study we built up an experimental setup in which conversational partners held a conversation once with and once without an optical feedback channel. A detailed transcription of the recorded speech material was segmented into turns. In each turn the topic units were identified and the syllables were labelled. We measured and compared prosodic feature characteristics between turn-final and turn-medial topic units.

1 Introduction

In a verbal conversation the roles of speaker and listener have to be defined. Sacks et al. (1974) stated “minimize gap and overlap” as the first rule for a working turn-taking-mechanism. According to them, the end of turn has to be marked in some way. Since linguistic cues are rarely found, it is obvious that this marking has to be realized by prosodic features. This supposition was corroborated by the findings of Lehiste (1975), that listeners got the ability to identify the position of clauses within a turn, even if the clauses were represented in isolation. In the speaker’s turn several prosodic cues are presumed to indicate to the listener whether the speaker wants to keep or end the turn. At points with high speaker switch potential noticeable gestural and mimic cues can be found. It is unknown how important those non-verbal aspects are for the turn-taking indication. The main research question of the presented study is:

Do prosodic cues compensate if the optical feedback channel is missing in the verbal conversation?

2 Prosodic and non-verbal turn-taking indicators

Duncan (1972) sorted turn-taking signals by their function. He classified the signals as turn-yielding, turn-demanding (listener), attempt-suppressing, and back-channel-communication (listener response).

We focused on turn-yielding as those signals are easy to locate, and because most found prosodic and non-verbal cues belong to this class. Beattie (1981) and Oreström (1983) showed that a noticeable rising or falling movement of fundamental frequency acts as a prosodic turn-yielding cue. According to Oreström (1983), the final syllable of the turn is lengthened and sometimes the syllable frequency is increased. Duncan (1972) and Oreström (1983) documented a decrease of intensity at the end of a turn.

In addition, non-verbal cues for turn-yielding have been suggested. Kendon (1967) noticed that a speaker often doesn’t look at the listener during an utterance but does so at the end of the turn. An explanation is that at those points of the conversation visual contact is required. Exline (1965) discovered that participants in a conversation look at their dialogue partner more often while they are listening. Duncan (1972) found several non-verbal cues in the behaviour of a speaker as turn-yielding signals: Relaxation of a tensed hand-position, completion of a gesture, regression of the torso, and relaxation of the facial expression.

3 Data Retrieval

3.1 Experimental setup

In our experiment speakers held two conversations, both in two conditions: first with eye-

contact and then without. Speakers didn't know each other. The given task was to plan a party by seating guests on a map of the party location. For solving the task it was necessary that the conversational partners share their information.

3.2 Preparation of recorded speech material

There were four speakers in two bilateral conversations. During the first half of a conversation the speakers could see each other. After they had accomplished half of the task a screen foreclosed eye-contact. The recordings were transliterated into orthographic text by a phonetic expert. This detailed transliteration contains information like word fragments, hesitations, pauses, and vocal events like laughter. The transcribed text was then segmented into turns. In each turn the topic units were identified according to our definition:

- A topic unit can be considered as semantically and grammatically complete and
- there is no further division possible in grammatically and semantically complete units.

Table 1 shows the number of topic units we found in our material. The syllables were labelled and the F_0 -contours were determined by manual judgment.

Table 1: Numbers of topic units for each speaker.

	Speaker No.	Condition 1	Condition 2
Number of Topic Units in turn-final Position	1	25	30
	2	17	24
	3	23	24
	4	20	25
Number of Topic Units turn-medial Position	1	92	150
	2	59	94
	3	86	129
	4	79	95

3.3 Acoustic Measurements

In the analysis of the acoustic speech signal we focused on features that have been suspected as turn-yielding signals in former studies. Each end of a topic unit has the potential to be the end of the turn and to initiate a turn taking. We assumed that a speaker marks topic units in turn-final position compared to turn-medial topic units by prosodic differences and that those differences change if the optical feedback channel is missing.

We observed the following prosodic features:

- Speech rate (syllables per second)
- Average intensity across topic units

- Difference of intensity of final last three syllables and non-final last three syllable of topic units (in Hertz)
- Mean F_0 in topic units (in Hertz)
- Mean range of F_0 in topic units (in Hertz)
- Difference in duration between final and non-final syllables of topic units (in ms)
- Relative distribution of five different closing F_0 -contours in the topic units
- Characteristic F_0 -values of five different closing F_0 -contours (manual judgment)

4 Findings

We intended to examine whether the differences between turn-final and turn-medial topic units differ in the feature characteristics between the two conditions. Feature characteristics could differ in quality or quantity. For a variation in quantity the number of potential signals would increase or decrease between the two conditions. A variation in quality could only be analyzed if the potential signal appears in both conditions and occurs as an increase or decrease of the strength of the feature.

4.1 Duration

The mean syllable rate of final topic unit and non-final topic unit was compared. Our results indicate an increased syllable frequency at the end of the turn in condition 2. But there is no significant difference between the two conditions. One speaker even decreased syllable frequency in turn-final positions compared to non-turn-final positions.

4.2 Intensity

For the intensity we analyzed differences between

- the overall intensity of the topic units in final and in non-final position, and
- the internal reduction of intensity at the end of the topic units in final and non-final position (by comparing the last three syllables to the others).

The overall intensity of topic units in turn-final position is significantly decreased for two (of four) speakers in the condition with sight and for three speakers in the condition without sight. That is, there seems to be a signal function which

is used by one more speaker in condition 2. However, this is just a quantitative difference between the two conditions. For the speakers, using this potential signal in both conditions, there's no detectable qualitative variation in condition 2 (no enhanced difference between the intensity of topic units in final and non-final position).

For all topic units a decrease of intensity at the end has been found. Due to the decrease of air pressure during an utterance this was expected. This reduction of intensity is for all four speakers only significant for topic units in final position. That is, that in topic units at the end of a turn the final reduction of intensity is much greater than in the other topic units. One could assume a signal function. Further analyses showed that this distinction is intensified by two of the speakers in condition 2, while it is weaker for the other two speakers. The modifications in condition 2 don't have a mutual direction.

4.3 Fundamental frequency

Concerning the fundamental frequency, we examined the following issues by comparing the two conditions:

- The over-all F₀-mean and F₀-range of the topic units in final and non-final position
- The percentage distribution of final F₀-contours at the end of the topic units in final and non-final position
- The representative last F₀-values of these contours in final and non-final position (last level tone for movements and F₀-mean for sustained F₀).

The speakers (exception is one speaker in condition 2) realized the turn-final topic units with lower fundamental frequency; which is significant only for two speakers in both conditions. These two speakers made a stronger distinction between final and non-final topic units in condition 2 by increasing the difference of mean F₀. The other two speakers diminish this distinction in condition 2.

Equivalent is the finding for the F₀-range. The same two speakers who decreased the mean frequency decreased also the F₀-range in final topic units. For these speakers there's also a noticeable intensification of the distinction in condition 2, while the other speakers behave contrarily.

Analyzing the percentage distribution of F₀-movements, we distinguished five F₀-contours at

the end of the topic units: *Sustain*, *Fall*, *Rise*, *Rise/fall*, and *Fall/rise*.

None of these contours seemed to appear more often in turn-final position. This includes the falling and rising F₀-movements, which were assumed to be turn-yielding signals. In contrast most of the topic units were realized with a final *sustain* and there was no higher occurrence of *rise* or *fall* in turn-final position detectable.

For the final level tone (F₀-mean for *sustain*) we found only for *sustain*, *fall* and *rise/fall* differences between final and non-final topic units. These contours had lower final level tones (lower F₀-mean for *sustain*) in turn-final position for at least one speaker in condition 1 and 2 speakers in condition 2. These findings accounted for the general lowering of F₀ in final topic units. Although this distinction between final and non-final position doesn't change qualitatively between conditions, there's some evidence for a quantitative change, because more speakers seem to use these signals in condition 2.

5 Discussion and Conclusion

To examine whether turn-yielding signals are intensified in the condition without sight, we constituted the criterion that a cue has to appear in one of the conditions for at least three of the speakers to be considered. For those cues we developed a comparison chart in which qualitative and quantitative changes between condition 1 and 2 (with and without sight) were inscribed (Table 2).

Qualitative differences could only occur if a speaker shows the signal in both conditions. They are treated dichotomous (as increased and decreased). Quantitative differences are marked as added or omitted signals for each speaker.

Table 2 shows that none of the signals undergoes changes of the same direction for more than two speakers. For syllable frequency there's no change between the conditions at all. For intensity of topic units, F₀-mean of *sustain*, and last level-tone of *fall* there's only a quantitative change for one speaker. That is, one more speaker added this signal in condition 2, while one or more speaker use it in both conditions, without qualitative shades.

Only the last level-tone of *rise/fall* and the difference of intensity between final and non-final syllable groups show changes for more than one speaker. But while the difference of intensity between final and non-final syllable groups is increased in condition 2 for the conversation

partners in group 1, it is decreased for group 2. The results cancel each other.

Only the last level-tone of *rise/fall* was modified in condition 2 by more than one speaker without being modified by other speakers in the contrary way. That is, it was added. For the mean F_0 and the F_0 -range one speaker omitted the signals in condition 2 and one speaker increased their distinctive function.

Taking a look at the sum of shown signals, we recognize that for none of the speakers there's a

remarkable raise in the total number of shown signals in condition 2. Finally, every increased distinctive function of a signal, which could be judged as compensation, has a negative counterpart like decrease or omission. Based on the results of this study one can't assume, that the analyzed prosodic cues compensate if the optical feedback channel is missing. This leaves the question whether optical cues are necessary signals or just added as redundant indicators to intensify the effect of the prosodic cues.

Table 2: Main results of the analysis of some analyzed prosodic features.

	Comparison of feature characteristics between turn-final and turn-medial topic units											
	Condition 1: with sight				Condition 2: without sight				Change from Condition 1 to 2			
	Conversation 1		Conversation 2		Conversation 1		Conversation 2		Conversation 1		Conversation 2	
	TP	MB	SK	SI	TP	MB	SK	SI	TP	MB	SK	SI
Syllable frequency	+	-	+	+	+	-	+	+
Intensity	-	+	-	+	+	+	-	+	++	.	.	.
Difference of intensity between final and non-final last three syllables	+	+	+	+	+	+	+	+	i	i	d	d
Mean fundamental frequency	+	+	+	-	+	+	-	-	.	i	--	.
Mean range of fundamental frequency	+	+	+	-	+	+	-	-	.	i	--	.
Mean fundamental frequency <i>sus</i>	+	-	-	-	+	-	+	-	.	.	++	.
Last level tone <i>fall</i>	+	-	-	-	+	-	+	-	.	.	++	.
Last level tone <i>rise/fall</i>	-	-	-	+	+	+	-	+	++	++	.	.
Sum	7	5	4	5	8	5	5	5				

+/-	indicator for signal function found/not found
.	no difference found
++	signal function added
--	signal function omitted
i/d	signal function increasing/decreasing

Acknowledgement

We thank the test speakers for volunteering.

References

- Adam Kendon. 1967. *Some functions of gaze-direction in social interaction*. Acta Psychologica 26: 22-63.
- Bengt Oreström. 1983. *Turn-taking in English Conversation*. Lund studies in english 66, Infotryck AB, Malmö.
- Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. *A simplest systematics for the organisation for turn-taking for conversations*. Language 50: 396-735.
- Ilse Lehiste. 1975. *The phonetic structure of paragraphs*. Structure and process in speech perception. Proceedings of the symposium of dynamic aspects

of speech perception. Cohen / Nooteboom (Eds.), Springer Berlin / Heidelberg, 195-206.

- Ralph V. Exline, David Gray, and Dorothy Schuette. 1965. *Visual behavior in a dyad as affected by interview content and sex of respondent*. Journal of Personality and Social Psychology 1: 201-209.
- Starkey Duncan, Jr.. 1972. *Some signals and rules for taking speaking turns in conversations*. Journal of Personality and Social Psychology 23: 283-292.
- W. Geoffrey Beattie. 1981. The regulation of speaker turns in face-to-face conversation: Some implications for conversation in sound-only communication channels. Semiotica 34: 55-70.