Looking for melodic turn-holding configurations in Dutch*

Johanneke Caspers
Universiteit Leiden/HIL.

1. Introduction

The generally smooth and fast alternation of speaking turns in every-day conversation is a phenomenon that may be explained only in terms of a highly complex system of interacting factors, comprising syntax, semantics, pragmatics, prosody, visual cues etc. However, before the intricacies of the turn-taking system can be determined, attention has to be given to the relevant subparts. In spite of the fact that the relevance of intonation for turn-taking is widely recognized, our knowledge of its precise function in the structuring of conversation is still fairly limited. My particular interest is in the role intonation may play in signaling the wish of the speaker to keep the turn.

The course of speaking turns in natural conversation must be guided by a system of rules adhered to by the interlocutors, otherwise the process would not run as swiftly and smoothly as it generally does. The well-known system proposed by Sacks, Schegloff and Jefferson (1974) has basically remained unchallenged (Levelt 1989). Starting from this basic analysis, a large number of empirical studies of natural conversation have been carried out, and several recent investigations display a specific interest in the role of speech melody (e.g. Auer 1996; Ford and Thompson 1996; Couper-Kuhlen 1996; Selting 1996; Wells and Macfarlane 1998). A characteristic of these analyses is that they are qualitative in nature, which means that no frequency counts are given for specific variables, and that evidence is presented in the form of detailed examples, annotated in an elaborate transcription system, sometimes complemented with simple instrumental analyses such as pause duration measurements.

A notion central to Conversation Analysis is the ‘Transition Relevant Place’ (TRP), a point of possible completion of the present turn, where turn-taking becomes relevant (but not obligatory). Interlocutors project these points of possible
turn completion before their actual occurrence — which explains the often jointless transition between consecutive speakers — and I would like to know more about the role that intonation plays in this projection of upcoming turn endings.

There is some evidence from the analysis of natural dialogue that pitch accents play a significant role in the projection of possible turn endings (Selting 1996; Wells and Macfarlane 1998). It seems that the shape of a potentially final pitch accent (plus the non-prominent tones that follow) can signal whether the current speaker wants to continue speaking or give the floor to someone else.

In earlier research, focusing on functional differences between two types of Dutch intonation contours (Caspers 1998), perception experiments were used in an attempt to establish the possible relevance of certain melodic structures for the turn-taking system. The two contours compared were uttered by trained speakers and comprised an accent-lending rise followed by level high pitch; they differed only in the presence versus absence of a final rise. Below examples are given, presenting stylized versions of the two investigated contours on a short utterance (a proper name), annotated in the ToDI transcription system (Gussenhoven, Rietveld and Terken 1999):

(1) Jan-WILlem
    H* %

(2) Jan-WILlem
    H* H%
    ‘Jan-Willem.’

In both cases there is an abrupt rise of the pitch on the syllable bearing the sentence accent ('WIL-'), while the pitch remains high on the syllables following the main accent. In (1) the contour ends without a specific boundary-marking pitch movement (which is represented by the label '%'), but in contour (2) there is a further rise on the final syllable, a so-called high boundary tone (indicated by 'H%'). The following hypothesis was tested: the contour ending in level high pitch (1) cannot readily be used to signal a question, as opposed to (2), but rather functions as a turn-holding cue. The contours were presented in controlled environments, in which listeners were asked to perform a number of comparison and rating tasks. The results of the experiments were clear: subjects indeed indicated that H* % cannot be interpreted as a signal that the listener should take the next turn. Instead, use of this contour was understood as a sign that the current speaker wanted to keep the turn (for more details see Caspers 1998).

These results led to the following research questions:

– does H* % (i.e., an accent-lending rise followed by level high pitch) function as a turn-holding cue in natural Dutch dialogue?
– if $H^*$ does function as a turn-holding cue, are there other melodic configurations with the same function?

2. Approach

In a first attempt to answer the research questions it was decided to investigate a corpus of Dutch production data. A quantitative analysis of the melodic and turn-transitional structures of natural conversations should provide some insight into the role that intonation plays in the turn-taking system. In order to find out whether a systematic relationship exists between speech melody and turn-taking, we need to locate possible turn transition places in the dialogue, and then examine these places for the transition type actually occurring and the melodic characteristics immediately preceding these positions.

I adopt the approach taken by Koiso, Horiuchi, Tutiya, Ichikawa and Den (1998) who use so-called Inter Pausal Units (IPUs) to divide spontaneous dialogues into turn-relevant components (see §4.1). IPUs can be determined objectively (even automatically) and the boundaries between these units can be labeled as instances of either turn-holding or turn-changing (see §4.2). The accent-lending and boundary-marking configurations preceding every IPU boundary were described in terms of the ToDI model (§4.3). After these labeling stages, the available data were viewed in the light of the research questions posed.

3. Corpus Materials

Use was made of a corpus of Dutch guided spontaneous conversations, so-called Map Task dialogues (cf. Anderson, Bader, Gurman Bard, Boyle, Doherty, Garrod, Isard, Kowtko, McAllister, Miller, Sotillo and Thompson 1991). In these task-oriented dialogues, maps provide a handle on an essentially spontaneous conversation. There are two roles in the Map Task dialogues: one of the interlocutors is the ‘instruction giver’ and the other participant is the ‘instruction follower’. The instruction giver has a map with a route drawn on it and his or her task is to explain to the instruction follower which route to draw on his or her unmarked copy of the map. The participants cannot see the map of the other party. Both maps have a number of reference points on it (e.g., ‘old pond’, ‘new pond’, ‘green meadow’) and by introducing small differences between these reference points it is possible to complicate the dialogue to some extent. For example, in the maps used for the present investigation (designed by Bob Ladd and Astrid Schepman), there were differences in presence and absence of some of the reference points (e.g., both interlocutors have a ‘new pond’, but only one has a ‘green meadow’), in the hope of eliciting questions and maybe even interruptions in the dialogues.
4. Analysis

For the acoustic analysis of the materials and storage of the various labels, the program Praat (Boersma 1999) was used. There were three labeling stages, which will be discussed in the following subsections.

4.1 Inter Pausal Units

The materials were divided into so-called Inter Pausal Units. An IPU is defined as “a stretch of a single speaker’s speech bounded by pauses longer than 100 ms” (Koiso et al. 1998:299). This means that boundaries were drawn in all positions where a pause longer than 100 ms appeared in the signal, and in positions where a change of speaker occurred. In this way, pauses longer than 100 ms were used as an objective criterion in the division of the dialogue into units that may be relevant to the turn-taking system.

A problem with this type of analysis is that other turn-relevant places could be missed, e.g., pauses shorter than 100 ms, and/or places where substantial lengthening occurs.

4.2 Transition types

The second stage of the analysis comprised a characterization of all IPU boundaries in terms of speaker transitions. Following Koiso et al. (1998), a division was made between the categories hold and change: the former meaning that the same speaker continues after a (turn-internal) pause of 100 ms or more, the latter indicating that a turn change has occurred (with or without an intervening pause). But there are more possibilities. For instance, Koiso et al. pay special attention to so-called ‘backchannels’, short optional utterances produced by the hearer to signal that s/he is still engaged in the dialogue, prompting the current speaker to continue. For instance, “yes” can be used to indicate that the listener has understood and that the speaker may continue; however, if the “yes” is an answer to a yes–no question it is not optional and therefore not a backchannel. As the name indicates, backchannels are not viewed as speaker turns, but as sounds occurring during the turn of another speaker. The distinction between backchannels and ‘real’ turns (change) is essential to the questions posed in the present paper.

While Koiso et al. leave overlapping speech outside their analysis, with the argument that they want to investigate only smooth transitions, the current analysis also tries to incorporate stretches of concurrent speech. Therefore, two additional transition types were defined: simultaneous start and interruption. In the case of a simultaneous start, both interlocutors start speaking at the same time, whereas
Looking for melodic turn-holding configurations in Dutch

An interruption is defined as a turn-competitive incoming (cf. Wells and Macfarlane 1998) of a participant during the turn of the current speaker. The five possible transition types are illustrated below; the boxes following S1 depict stretches of speech (without pauses larger than 100 ms) uttered by speaker 1, S2 indicates the speech by speaker 2, the dotted line marks the time course, and the arrow indicates the relevant IPU boundary. Note that the inclusion of overlapping stretches of speech in the analysis leads to somewhat broader definitions of the transitions types change and backchannel than those used by Koiso et al. (1998).

The definition of the transition type backchannel is somewhat problematic because it is not always easy to determine whether an utterance is optional or not. Another issue is the line to be drawn between change and interruption. In normal conversation, partial overlap between consecutive turns is a common phenomenon, which can be explained by the notion of Transition Relevant Place (e.g., Sacks et al. 1974; Wells and Macfarlane 1998), since one of the characteristics of a TRP is that it may appear some distance before the actual end of the utterance. Therefore, no rule is violated when an interlocutor takes the turn at, or just after, a TRP that is situated before the point of completion of the current turn. Thus, overlap in speech does not necessarily mean that an interruption has occurred. However, when a speaker comes in before the TRP, it can be viewed as turn-competitive and therefore as an interruption (cf. Wells and Macfarlane 1998). In practice it is not often a problem to distinguish change from interruption, even if the notion of TRP is hard to define; in most cases it is clear whether some sort of
syntactic completion has occurred or not, which coincides with the perception of a normal speaker change or an interruption, respectively.

4.3 Melodic characteristics

The third stage of the analysis comprised a transcription of the melodic phenomena occurring immediately before each IPU boundary. As a tool for labeling the melodic phenomena, the recently developed ToDI ('Transcription of Dutch Intonation', Gussenhoven et al. 1999) system was chosen. Starting from the principle that the “potential last accent” (cf. Selting 1996) is of major importance to the turn-taking system, the last pitch accent before every IPU boundary was transcribed, as well as the tone sequence following this accent up to the boundary. The intonation was labeled on the basis of the auditory impression of the pitch curve only. Before every IPU boundary a boundary tone was transcribed, which means that intonation domain boundaries were determined by pauses or speaker changes actually occurring in the material, and not by the syntactic structure of the utterance. Note that, as a result, the boundary tones marked in the current analysis do not necessarily correspond to the boundary tones as defined by ToDI.

One of the problems in this stage was the transcription of the overlapping stretches of speech. In most cases these segments could not be labeled, but this did not have a large influence on the analysis as a whole. Substantial overlap occurs when the two interlocutors take a turn at the same moment (SIMULTANEOUS START) or when one interlocutor interrupts the other (INTERRUPTION); in general these competitive turns take only a limited amount of time, leaving one of the interlocutors as the ‘winner’. While the ‘loser’ drops out, the winning speaker simply continues the turn without any obvious break. Therefore, the final pitch accent plus whatever follows up to the next IPU boundary can be labeled normally.

Another issue was the transcription of the melodic configuration found on the majority of the backchannels: a clear drop in pitch immediately followed by a rise, without the suggestion of prominence, i.e., without a clear pitch accent. ToDI does not offer a suitable label (neither does the Grammar of Dutch Intonation, cf. ’t Hart, Collier and Cohen 1990); therefore I decided to give these specific configurations the label LH% (a low tone followed by a high boundary tone). Furthermore, the label M (for ‘mid’) was used to transcribe the level mid tone of so-called ‘filled pauses’, predominantly vocalized hesitations like “uhm”.

4.4 An example

To illustrate the analysis described above, Figure 1 presents a small segment of labeled dialogue. The upper part of the figure shows the \( F_0 \) curve, giving an impression of the course of the pitch during the entire segment (straight lines
Looking for melodic turn-holding configurations in Dutch

Figure 1. Example. Above: $F_0$ curve (in Hz); below: waveform plus labeled tiers, pause = length pauses > 0.1 s, S1 = text instruction giver, S2 = text instruction follower, gloss = verbatim translation, trans. = transition type (H = hold, S = simultaneous start, C = change, B = backchannel), ToDi = melodic labels.
represent the voiceless parts). The lower part of the figure shows the waveform and six labeled tiers: the upper tier gives the position and length (in seconds) of pauses longer than 100 ms ($\S 4.1$); the second one presents an orthographic transcription of the speech uttered by the (male) instruction giver (capitals mark the final pitch accent before each IPU boundary) and the third tier offers the speech by the (female) instruction follower, while the fourth tier gives a verbatim translation; the fifth tier contains the transition types ($\S 4.2$) and the lowest tier presents the ToDI transcription of the melodic configurations immediately before each break ($\S 4.3$).

4.5 Hypothesis

Six complete Map Task dialogues were analyzed, which amounts to a little over 30 minutes of speech. In total, 1,070 IPU boundaries were annotated for transition type and melodic type. It was hypothesized that an accent-lending rise followed by a level high tone ($H^* %$) would predominantly be followed by either hold, backchannel or interruption. These three transition types can be viewed as instances of turn-holding, since backchannels occur by definition within the turn of another speaker (and are a prompt for the current speaker to continue) and interruptions happen at positions where the original speaker intended to continue, whereas both change and simultaneous start involve a change of speaker.

5. Results

Table 1 contains the frequency of occurrence of the various transition types encountered after all instances of $H^* %$ found in the material ($N=90$). The table shows that in more than half of the cases the configuration $H^* %$ is followed by an utterance by the same speaker (hold), in 12% by a backchannel and in 16% by an interruption, adding up to a total of 85% in favor of the hypothesis concerning the

<table>
<thead>
<tr>
<th>Transition type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>hold</td>
<td>51 (57%)</td>
</tr>
<tr>
<td>backchannel</td>
<td>11 (12%)</td>
</tr>
<tr>
<td>interruption</td>
<td>14 (16%)</td>
</tr>
<tr>
<td>change</td>
<td>10 (11%)</td>
</tr>
<tr>
<td>simultaneous start</td>
<td>4 (4%)</td>
</tr>
<tr>
<td>total</td>
<td>90 (100%)</td>
</tr>
</tbody>
</table>
Looking for melodic turn-holding configurations in Dutch

The amount of available data is relatively small, since only 90 of the 1,070 IPU boundaries analyzed were preceded by H* % (8%).

The fact that there is no clear boundary tone in the H* % configuration — the pitch remains flat until the following IPU boundary and does not drop to a clear low point (L%) or rise to a high point (H%) on the final syllable — may be crucial to the issue of turn-taking. In the ToDI system, nuclear contours ending in % are described as “half-completed”, clearly pointing to the unfinished nature of such melodic configurations. The following question presented itself: do all configurations ending in % generally function as turn-holding cues?

In Table 2 the frequency of transitions involving turn-changing (types change and simultaneous start) versus the transitions involving turn-holding (hold, backchannel and interruption) is presented, broken down by boundary tone type. When an IPU boundary is preceded by a low boundary tone (L%) the turn is changed in 55% of the cases; after a high boundary tone (H%) there is a speaker change in 48% of the cases. This means that for low as well as high boundary tones the chance for a turn change is roughly 50%. However, when there is no clear boundary tone (%), the turn remains with the same speaker in 84% of the cases.

<table>
<thead>
<tr>
<th>boundary tone</th>
<th>+change</th>
<th>−change</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>L%</td>
<td>213 (55%)</td>
<td>177 (45%)</td>
<td>390 (100%)</td>
</tr>
<tr>
<td>H%</td>
<td>212 (48%)</td>
<td>234 (52%)</td>
<td>446 (100%)</td>
</tr>
<tr>
<td>%</td>
<td>37 (16%)</td>
<td>197 (84%)</td>
<td>234 (100%)</td>
</tr>
<tr>
<td>total</td>
<td>462 (43%)</td>
<td>608 (57%)</td>
<td>1070 (100%)</td>
</tr>
</tbody>
</table>

6. Conclusion and discussion

The data presented in Section 5 provide support for the hypothesis that an accent-lending rise followed by level pitch (H* %) functions as a turn-keeping cue in Dutch spontaneous dialogues. More generally, the data show that when there is no clear boundary tone immediately before an IPU boundary, there is no subsequent speaker change in approximately 85% of the cases, whereas a low or high boundary tone is followed by a change of speaker in approximately 50% of the cases. These results indicate that the absence of a clear boundary tone is generally taken as a sign that the current speaker has not finished yet (and wants to continue). This tentative conclusion has to be supported by more data, other types of analyses and perception experiments.
The striking absence of an effect of the low and high boundary tones on the occurrence of a speaker change (Table 2) demonstrates the relevance of other factors besides intonation for the turn-taking system. Previous work (see Caspers 1998 and the references given there) already showed that a melodic configuration like H* H% (see example (2) in §1) is ambiguous between signaling a question — prompting the listener to take the next turn — and something like a comma, i.e., signaling the intention of the current speaker to continue speaking. Participants in a conversation generally know which of the two possibilities is intended from other sources of information. The fact that there is some evidence for the existence of a melodic turn-holding cue, but not for specific turn-yielding cues, is in line with the work of Selting (1996). Larger data sets are necessary to establish whether the final pitch accent type also influences the odds for a change of turn; the present dataset is too small to provide a solid answer to this question. However, it seems that an H* pitch accent is more likely to be followed by turn-keeping than an H*L or an L*H pitch accent, whereas there is no pitch accent type that is typically followed by a change of turn. In their paper presenting a qualitative analysis of British English conversation Wells and Macfarlane (1998) conclude that there are specific “TRP-projecting accents” (pitch accents that project a Transition Relevant Place) as well as non-TRP-projecting accents, and that both types of accent can be defined on independent phonetic grounds. Further quantitative research is needed to determine whether English differs from Dutch and German in the presence versus absence of specific melodic configurations that project an upcoming change of speaker, or whether English exploits melodic turn-holding devices but no unique melodic turn-yielding cues, as seems to be the case in Dutch and German.

Notes
* This research was funded by the Netherlands Organisation for Scientific Research (NWO) under project #355–75–002. Vincent van Heuven is thanked for his valuable comments.
1. The fact that overlap sometimes accompanies turn transitions that can be viewed as perfectly ‘smooth’ may be interpreted as an argument against the choice by Koiso et al. (1998:301) to mark all overlapping transitions as ‘nonsmooth’; however, the Japanese turn-taking system may differ from the one employed in the Germanic languages in this respect.

References


