Melodic cues to turn-taking in English: evidence from perception

Anne WICHMANN
Department of Cultural Studies,
University of Central Lancashire
Preston PR1 2HE,
United Kingdom,
awichmann@uclan.ac.uk

Johanneke CASPERS
Phonetics Laboratory,
Universiteit Leiden Centre for Linguistics
Cleveringaplaats 1
2311 BD Leiden, The Netherlands,
j.caspers@let.leidenuniv.nl

Abstract

This paper presents a study of the effects of syntax and melodic configuration on turn-taking in Southern British English. Using dialogue materials, two perception experiments were carried out. In the first, subjects heard dialogue fragments in which syntactic completeness and melodic contour were systematically varied, and were asked whether they expected a subsequent turn exchange or not. In the second, subjects were presented with short speaker exchanges taken from the same material, and asked whether they thought the first speaker had intended to cede the turn or not. The results suggest that syntactic completion or non-completion is the main factor in predicting turn-taking behaviour. Only one melodic contour, the high level tone H* %, appears to operate as a turn holding device, regardless of whether the utterance is grammatically complete or not. The results of this study were found to be similar to those of a study of Dutch turn-taking.

Introduction

Most studies of the intonational cues to turn-taking have been carried out qualitatively within the theoretical framework of Conversation Analysis (e.g. Wells & Macfarlane 1998, Selting 1996). An exception to this is the study by Ford and Thompson (1996), who found that turn-changes in American English mostly appear when melodic, syntactic and pragmatic completion coincide. Two recent studies of the melodic cues to turn-taking in Dutch (Caspers 2000, 2001) motivate the present study, which uses comparable English data to replicate as far as possible the perception experiments carried out in Caspers (2001). On the basis of the findings for Dutch we expected syntactic completeness to be the overriding predictor of a possible turn change. Where melody has an effect, we hypothesised that, as in Dutch, the high level tone was likely to signal more to come and that no subsequent turn change would be expected (Caspers 1998). We also hoped to gain some insight into possible similarities and differences between the two languages.

1 Materials

The data used for these experiments was taken from Map Task data, recorded according to the Map Task conventions described in Anderson et al. (1991), and collected as part of the IViE project (Grabe et al. in preparation). We chose dialogues recorded in Cambridge, as they represented most closely the standard southern variety of British English.

2 Data Analysis

Following the method used by Caspers (2001), two complete Map Tasks (approximately 30 minutes of speech) were divided into inter-pausal units (IPUs, cf. Koiso et al. 1998) using a pause threshold of 100ms. Boundaries were then categorised according to three criteria.

Firstly, each IPU boundary was identified as either occurring within the turn of the same speaker, (category HOLD) or involving a change of speaker (category CHANGE). Secondly, the text before each IPU boundary was judged for syntactic completion. If the utterance was at least potentially complete at that point the boundary was categorised as syntactically complete, otherwise as 'not complete'. Finally, we selected a set of IPUs according to their final contour, identified in terms of the British system of 'nuclear tones' (fall, rise etc.) and in

autosegmental-metrical terms of final pitch accent and subsequent boundary tone. (Following Gussenhoven et al. (1999) we included the possibility of a boundary tone that was neither low nor high, transcribed as %.) We identified IPUs ending in one of the following contours: a high rise (H* H%), a high level (H* %), a fall-rise (H*L H%), a fall (H*L L%) and a truncated fall (H*L %).

3 Method

In the first experiment, subjects were presented with a dialogue fragment ending at an IPU boundary. The subjects' task was to predict what happens next, i.e. whether the speaker holds the turn, or holds after a brief backchannel response, or cedes the turn. In the second experiment we again asked subjects to judge what the first speaker had intended - to continue or to cede the turn, but under slightly different conditions: this time subjects heard brief exchanges involving both speakers. This was to see if the presence of an actual response influenced subjects' judgement of the first speaker's intention. The same subjects took part in both experiments. They were 25 native speakers of Southern British English, 9 men and 16 women, aged between 19 and 54, only 7 of whom had some background in linguistics. No hearing difficulties were reported.

4 Experiment One

4.1 Stimulus material

For the first experiment, the stimuli consisted of dialogue fragments, around 8 to 13 seconds in length, and ending in an IPU. The fragments were chosen such that they ended according to the following four conditions:

(i) turn exchange plus syntactic completion
(ii) turn exchange minus syntactic completion
(iii) turn hold plus syntactic completion
(iv) turn hold minus syntactic completion

The five contours chosen were as listed in paragraph 2 above. For all but the high rise (H* H%) two stimuli were chosen for each of the above conditions, giving 32 stimuli. As syntactic completion could, of course, include interrogatives, which would be highly likely to project a turn change, these were avoided for all but one stimulus for the fall-rise contour and three for the high rise. We found very few cases of the high rise in the English data, an interesting finding in itself, and it was not possible to find examples for each condition; only six cases were used altogether, four syntactically complete (two interrogatives and two declaratives) and two syntactically incomplete.

4.2 Procedure

After three practice examples, the 38 randomised stimuli were each presented twice. Subjects were asked to predict whether (1) the current speaker would continue, (2) the current speaker would continue after a short, non-obligatory backchannel response, or (3) the second speaker would take over.

4.3 Results

The results for this experiment are given in Tables 1 and 2. Table 1 gives the frequency of responses per condition, and Table 2 shows whether the differences in number of turn-keeping responses between the contour types are significant. Note that in the latter table we conflate the responses 'hold' and backchannel', since, despite subtle pragmatic differences, we judged the prediction of a backchannel to entail the prediction of a turn continuation (cf. Koiso et al.). A hierarchical loglinear analysis performed on the factors response type, contour type and syntactic completion shows significant associations between response type and contour type (partial χ^2=288.3, p<.0001), between syntactic completion and response type (partial χ^2=288.3, p<.0001), and interaction between the three factors (Pearson χ^2=507.4, p<.0001). This means that there are main effects as well as interaction effects of contour type and grammatical completion on the responses.

Table 1 shows that subjects virtually never expect a turn change when the fragment is syntactically incomplete (2%). The only significant differences in the number of expected turn-keepings ('backchannel' plus 'hold') are found between contours H*L L% and H*L H% and between H*L L% and H*L %, but these effects are rather small (see Table 2). The main difference appears to be the degree to which contours invite a backchannel response. This

Table 1. Part A; absolute (and relative) frequency of expected transition type ('change', 'backchannel' or 'hold') per contour type, broken down by syntactic completion ('minus' or 'plus').

	minus syntactic completion			
Contour	change	backchannel	hold	Total
H* %	1 (1%)	6 (6%)	93 (93%)	100
H* H%	1 (2%)	18 (36%)	31 (62%)	50
H*L L%	7 (7%)	19 (19%)	74 (74%)	100
H*L H%	1 (1%)	49 (49%)	50 (50%)	100
H*L %	-	14 (14%)	86 (86%)	100
total	10 (2%)	106 (24%)	334 (74%)	450
	plus syntactic completion			
contour	change	backchannel	hold	total
H* %	5 (5%)	6 (6%)	89 (89%)	100
H* H%	51 (51%)	48 (48%)	1 (1%)	100
H*L L%	59 (59%)	27 (27%)	14 (14%)	100
H*L H%	44 (44%)	38 (38%)	18 (18%)	100
H*L %	29 (29%)	31 (31%)	40 (40%)	100
total	188 (38%)	150 (30%)	162 (32%)	500

*Table 2. Part A; values of partial χ^2 tests (Pearson) on the turn-keeping responses (backchannel plus hold) for all pairs of contour types, broken down by syntactic completion; * indicates p<.05.*

	Minus syntactic completion			
contour	H* %	H* H%	H*L L%	H*L H%
H* H%	0.3			
H*L L%	4.7	1.7		
H*L H%	0.0	0.3	4.7*	
H*L %	1.0	2.0	7.3*	1.0
	plus syntactic completion			
contour	H* %	H* H%	H*L L%	H*L H%
H* H%	52.5*			
H*L L%	67.0*	1.3		
H*L H%	41.1*	1.0	4.5*	
H*L %	20.4*	10.1*	18.3*	4.9*

tendency is weak for both the fall (H*L L%) and the truncated fall (H*L %), but nearly half of the H*L H% contours in syntactically incomplete positions are judged to invite backchannel feedback.

The syntactically complete utterances, on the other hand, show a clear effect of contour type: a rising pitch accent followed by a level boundary tone (H* %) leads to 89% expected 'hold' responses, supporting the hypothesis that this melodic configuration functions as a turn-keeping device. In this respect it differs strongly from all other contours, as is evident from the data presented in Table 2.

The results for the syntactically complete H* H% stimuli reflect the utterance type, and should therefore be treated with caution. Not surprisingly, the two interrogatives attracted almost exclusively the judgement 'change'; the remaining two declaratives attracted almost exclusively the judgement 'backchannel'. The use of a high rise on declaratives is a recent and highly marked innovation in British English, and is assumed to have the function of eliciting hearer acknowledgment. Our results are consistent with this view.

As Table 2 shows, there was an interesting and significant difference between the effect of the complete fall (H*L L%) and the truncated fall (H*L %). The truncated fall is much more likely to cue a turn hold (71% of responses compared with 41% for the complete fall) and

correspondingly less likely to cue a turn change (29% compared with 59% for the complete fall).

4.4 Discussion

The results of this experiment suggest that, in this variety of English, incomplete syntax overrides any melodic cues. Only the high level tone appears to be a strong turn keeping device, regardless of syntax. On the other hand there appear to be no melodic contours which, when they occur in conjunction with syntactic completeness, can be said to predict a turn change. We thus find more evidence for the use of melody as a turn keeping device than as a turn ceding device. The second experiment was designed to investigate the degree to which such judgements of speaker intention were upheld in the presence of an actual speaker response.

5 Experiment Two

5.1 Stimulus material

The stimuli for this part of the experiment were drawn from the same material as in Part A. Each fragment that ended in the original data in a turn exchange was extended to include the turn exchange itself. This produced a sound file of around 8 to 12 seconds in length. The turn exchange was then excised as a short separate file of around 3 to 5 seconds. Regardless of contour, a speaker change at a syntactically incomplete point was hard to find in our data, and a number of these stimuli were created artificially by editing out intervening material.

5.2 Procedure

The same subjects participated in both parts of the experiment. They were first presented with the longer fragment containing the relevant turn exchange, and then heard the file containing only the turn exchange twice in succession. The 20 stimuli (4 for each contour) were preceded by three test stimuli. The subjects were asked to judge whether the first speaker had expected the turn exchange, had expected to continue, or whether it was unclear.

5.3 Results

Tables 3 and 4 contain the results for the second experiment. A hierarchical loglinear analysis performed on the factors response type, contour type and syntactic completion shows significant associations between response type and contour type (partial χ^2=143.8, p<.0001), between syntactic completion and response type (partial χ^2=200.5, p<.0001), and interaction between the three factors (partial χ^2=282.6, p<.0001). Again the biggest effects of contour type are found for the syntactically complete points: subjects do not think the original speaker wanted to yield his

Table 3. Part B; absolute (and relative) frequency of judged speaker intention ('change', 'unclear or 'hold') per contour type, broken down by syntactic completion ('minus' or 'plus').

	minus syntactic completion			
contour	change	unclear	hold	total
H* %	3 (6%)	2 (4%)	45 (90%)	50
H* H%	1 (4%)	2 (8%)	22 (88%)	25
H*L L%	13 (26%)	9 (18%)	28 (56%)	50
H*L H%	15 (30%)	11 (22%)	24 (48%)	50
H*L %	1 (2%)	5 (10%)	44 (88%)	50
total	33 (15%)	29 (13%)	163 (72%)	225
	plus syntactic completion			
contour	change	unclear	hold	total
H* %	4 (8%)	9 (18%)	37 (74%)	50
H* H%	66 (88%)	3 (4%)	6 (8%)	75
H*L L%	35 (70%)	11 (22%)	4 (8%)	50
H*L H%	39 (78%)	8 (16%)	3 (6%)	50
H*L %	43 (86%)	5 (10%)	2 (4%)	50
total	187 (68%)	36 (13%)	52 (19%)	275

*Table 4. Part B; values of partial χ^2 tests (Pearson) on the turn-keeping responses (hold) for all pairs of contour types, broken down by syntactic completion; * indicates p<.05.*

	minus syntactic completion			
contour	H* %	H* H%	H*L L%	H*L H%
H* H%	0.1			
H*L L%	7.4*	5.3*		
H*L H%	9.8*	6.7*	0.2	
H*L %	1.0	0.3	12.0*	14.6*
	plus syntactic completion			
contour	H* %	H* H%	H*L L%	H*L H%
H* H%	77.9*			
H*L L%	40.4*	6.3*		
H*L H%	50.0*	2.2	0.8	
H*L %	61.1*	0.1	3.7	1.1

his or her turn after a high level contour (there are only 8% expected changes after H* %), and Table 4 shows large differences between this contour type and all others. In contrast with the first experiment, however, there is a clear influence of contour type on the responses in the minus syntactic completion condition: in almost a third of the cases subjects feel that the original speaker had expected the turn to change after a default pitch accent (H*L) followed by a low (L%) or high (H%) boundary tone, that is, after a complete fall or after a fall-rise, and Table 4 shows that these two contour types differ significantly from all others (except from each other). The similarity between the complete fall and the fall-rise, which is also evident in the syntactic completion condition, suggests that both contours are perceived to have a similar function with respect to turn-taking and to be at least strong secondary cues to turn completion. In cases where there is a clear mismatch between syntax and contour (i.e. melodic completion but no syntactic completion) the actual presence of a speaker change makes subjects more likely to judge that this was the intention of the first speaker than they were in the first experiment, where they did not know what happened next.

Although subjects were simply asked to judge what they thought the first speaker had intended, their judgements were probably to some extent based on a post hoc analysis of the whole exchange. It is a general principle of pragmatics that utterances will be assumed to be relevant unless proved otherwise, and that conversational interaction will be assumed to be cooperative unless proved otherwise. There is therefore a strong likelihood that subjects subconsciously sought a cooperative explanation for actual turn changes wherever possible.

6 General Discussion

The major finding of this study, especially of the first part, is that if an isolated utterance is syntactically incomplete, listeners are highly unlikely to predict a turn change, whatever the melodic contour used. Where the syntax is complete, none of the contours lead listeners to predict exclusively a turn change. This means that both hold and change are possible at this point. There is one exception, namely where the accompanying contour is a high level tone (H* %). This contour in English appears to signal a clear turn hold, regardless of syntax.

We were also able to make some cross-linguistic comparisons. First, the similarities: it appears that in both Southern British English and Dutch the H* % contour signals the speaker's intention to keep the turn. This effect cannot be attributed to the absence of a 'real' boundary tone, since the truncated fall, which also ends in a %, does not behave as a cue to turn-keeping.

We also observed two main differences between the languages. The first concerns the occurrence of high rise tones (H* H%): we had difficulty in finding any of these in the English

data but not in the Dutch, which may indicate a general difference in contour distribution, or a difference in contour function in the two languages. This is an interesting question to pose in a larger-scale, corpus-based study.

The second difference relates to our observation that some contours are more likely than others to suggest a subsequent backchannel response. This has important implications for the study of cooperation in interaction, both within and between languages (cf. Wichmann 2000). The number of 'backchannel' judgements given as responses to the stimuli ending in a high level tone H* % differs between Dutch and English: Caspers (2001) reports that in the Dutch study 56% of these contours suggest a backchannel response, compared to only 6% in the English study. This difference may have consequences for cross-cultural communication: if types of conversational behaviour are 'appropriate' in one language but not in the other there is potential for cross-cultural misunderstandings which may be perceived as 'attitudinal'.

7 Conclusion

The results of this study of English turn-taking support the Dutch findings of Caspers (2001), suggesting that while there are no melodic contours which reliably predict a turn change, the high level contour (H* %) creates the strong percept in both languages of a turn continuation, regardless of whether the utterance is syntactically complete or not. Other contours appear to operate at most as secondary cues to turn-taking, with syntactic completion or non-completion having the stronger effect.

A further observation - that some contours are more amenable to a backchannel response than others - suggests differences between the two languages which may have important cross-cultural implications.

While the answers to some of these questions may more suitably be sought using other methods, notably corpus-based analysis, we consider that such approaches are complementary to the perceptual evidence reported here.

Acknowledgements

Thanks to Brechtje Post for providing the data, Rachael-Anne Knight for helping with the experiments, and Bill Nelson and Geoffrey Potter for technical support. Wichmann was supported by the AHRB (Arts and Humanities Research Board, UK) research leave scheme; Caspers' work was supported by the Netherlands Organization for Scientific Research (NWO), under project #355-75-002.

References

A.H Anderson, M. Bader, E. Gurman Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo & H.S. Thompson (1991) 'The HCRC Map Task Corpus' *Language and Speech* 34, 351-366.

J. Caspers (1998) 'Who's next? The melodic marking of question vs continuation in Dutch.' *Language and Speech* 41, 375-398.

J. Caspers (2000) 'Looking for melodic turn-holding configurations in Dutch.' *Linguistics in the Netherlands 2000*, John Benjamins, Amsterdam, 45-55.

J. Caspers (2001) 'Testing the perceptual relevance of syntactic completion and melodic configuration for turn-taking in Dutch.' *Proceedings 7th European Conference on Speech Communication and Technology*, Aalborg, 1395-1398.

C.E. Ford and S.A. Thompson (1996) 'Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns.' In E. Ochs, E.A. Schegloff and S.A. Thompson (eds) *Interaction and Grammar*. Cambridge University Press, Cambridge, 134-184.

E. Grabe, B. Post and F. Nolan (in preparation) *Intonational Variation in the British Isles. Evidence from varieties of English spoken in Cambridge, Belfast and Bradford.*

`http://www.mml.cam.ac.uk/ling/ivyweb/intoproj.HTML`

C. Gussenhoven, T. Rietveld and J. Terken (1999) 'ToDI, Transcription of Dutch Intonation', `http://lands.let.kun.nl/todi`

H. Koiso, Y. Horiuchi, S. Tutiya, A. Ichikawa and Y. Den (1998) 'An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs'. *Language and Speech* 41, 295-321.

M. Selting (1996) 'On the interplay of syntax and prosody in the constitution of turn-constructional units in turns in conversation.' *Pragmatics* 6, 367-388.

B. Wells and S. Macfarlane (1998) 'Prosody as an interactional resource: turn projection and overlap.' *Language and Speech* 41, 265-294.

A. Wichmann (2000) *Intonation in Text and Discourse*. Pearson Education, London.