ON THE ANALYSIS OF PROSODIC TURN-TAKING CUES

Anne Cutler and Mark Pearson

1. *Introduction*

Would you mind just letting me finish?

Why can I never get a word in edgeways?

What's up? Cat got your tongue?

When a conversation breaks down, the problem can often be traced to a failure in the turn-taking procedure, i.e. the smooth interchange of speaking turns between conversational partners. For a conversation to function successfully, each speaker's turn should not go on too long, and should be accomplished without interruption; and at the end of one speaker's turn another speaker should take over without too long an intervening pause. Of course, at what point a turn of an inter-turn pause becomes "too long" may depend upon the particular conversational circumstances - e.g. on such factors as how well the participants know one another, their relative age or social status, and the difficulty of the subject matter under discussion. For any given conversation, however, it is usually obvious whether or not it is proceeding smoothly.

To take over the turn at the appropriate moment, without undue hesitation, it is obviously useful to be able to decide as early as possible that the previous speaker has finished or is about to fininsh. Clearly, syntax, semantics and reference to the discourse context play the largest role here. A completed utterance usually forms a syntactically complete unit. Questions usually signal that a response is required. Anecdotes have ends (if not always punch lines). And there are many more text-internal cues to whether or not a speaker has finished.

External to the text, however, there exists a considerable range of cues which speakers may employ - consciously or not - to inform hearers where the current turn will end. Some of these cues are paralinguistic in nature - i.e. not part of the speech signal at all. For example, speakers often look away from the interlocutors while speaking, but look towards them again as they finish talking (Kendon, 1967), especially if speaker and interlocutor do not know each other well (Rutter et al, 1978) and especially if the topic under discussion is difficult (Beattie, 1979). Termination of a hand gesture has also been claimed to be associated with turn-final utterances (Duncan, 1972). But there are also text-external cues which are part of the spoken utterance, and it is with these cues - specifically, those borne by the utterance prosody - that the present paper is concerned.

2. Prosodic Structure and Turn-Taking

The prosodic structure of speech comprises variation in three dimensions - fundamental frequency, duration and amplitude. All three dimensions exhibit specific effects which are dependent upon utterance position. Thus in the unmarked case fundamental frequency declines across the utterance (Maeda, 1976), as does amplitude (Lieberman, 1967); violations of these effects are marked as carrying information - for example, the terminal rise associated with certain question forms. Timing relations vary similarly; a given word will be uttered with longer duration in phrase-final than in non-phrase-final position (Oiler, 1973), although there seems to be no evidence that utterance-final lengthening is greater still (Oiler, 1973; Lehiste, 1980). Thus at the end of any utterance, whether or not it completes the speaker's turn, we would expect to find a fall in baseline pitch, a decrease in amplitude and some segmental lengthening. Prosodic turn-yielding cues, if any, would have to be overlaid upon this characteristic utterance-final prosodic pattern. Very closely related to this prosodic configuration, in addition, are certain voice quality features - e.g. creaky voice - which may also function as turn signals.

Duncan (e.g. 1972) has claimed that prosodic turn-ceding cues indeed exist. In a series of papers (Duncan, 1972,1973,1974,1975) he has reported a major study in which two 20-minute two-person conversational interactions were transcribed in detail from videotape. In this study Duncan identified six "turn signals", three of which were prosodic, namely:

(1) "The use of any pitch level/terminal juncture combination other than 22/ at the end of a phonemic clause" (22/here refers to a sustained "mid" pitch level in the Trager and Smith (1951) system).
(2) "Drawl on the final syllable or on the stressed syllable of a phonemic clause".
(3) "A drop in paralinguistic pitch and/or loudness in conjunction with [one of several stereotyped expressions such as "but uh", "or something" or "you know"]". (All quotations from Duncan, 1973, p.37).

With hand gesture termination, the stereotyped expressions referred to under (3) above, and completion of a syntactic clause (!), these cues are said to comprise the repertoire of potential turn signals. By labelling them in this way, however, Duncan is clearly begging the question; the term "signal" implies a communicative function between speaker and receiver which is in no sense justified by Duncan's analysis. Take, for example, the "signal" of syntactic clause completion. Clauses are completed frequently in speech, but only a small proportion of them also complete conversational turns. If clause completion were indeed an effective "turn signal" we would presumably find our interlocutors wanting to resume speaking, every time we finished any clause. As this does not generally happen - at least in the authors' experience - we can assume that the effectiveness of clause completion alone as a turn-yielding signal is in fact very slight.

It is not clear from Duncan's publications how he arrived at his particular set of signals, although it is implied that they were simply collated after inspection of the transcriptions of those utterances which ended speakers' conversational turns. A more proper term for the phenomena he listed would therefore be "correlates of end of speaking turn". Duncan used his compilation of "turn signals" to generate and test predictions about the relationship between number of "signals" produced at once and interlocutors' attempts to resume speaking at that point. This procedure is logically circular, and in this particular instance the results were also statistically very shaky (see the criticisms advanced by Beattie, 1981).

Moreover, Duncan's prosodic descriptions are extremely ill-defined. Signal (1) is described in terms of a particular Trager-Smith pitch level; but this kind of pitch notation is notoriously subject to influence from the syntax and semantics of the utterance (Lieberman, 1965). Furthermore, (1) is not even expressed as turn-yielding signal at all, but rather as what is *not* a turn-yielding signal - what it says in effect is that a sustained middle pitch (that is, presumably, neither a rise nor a fall) is a signal that the speaker wishes to *hold* the turn. Signal (2) is "drawl", which is not defined - although the term presumably refers to a phrase-final lengthening which is greater than would be expected in the default case, no metric is given for determining the relation between expected and observed phrase-final lengthening for any particular utterance. (This is a somewhat complicated procedure, but it is clear how it should be done; Lehiste (1980) gives an excellent and instructive example. She computed the average duration of every segment of a particular type - e.g. voiced stops, diphthongs etc. - in a stretch of speech, and then compared the actual length of words in phrase-final, paragraph-final and non-final positions in comparison with their expected lengths as computed by summing the average durations of their constituent sounds.) Finally, signal (3) - a drop in pitch and/or loudness - is, as we saw above, the default case for utterance-final phrases whether stereotyped or not. One suspects that here Duncan is actually talking about vocal quality features - creaky or whispery voice. But as with the other "turn signals" one retains the impression that Duncan merely recorded a subjective impression of what he heard. As all the speech in his study was apparently transcribed with full reference to the discourse context, there would have been considerable scope for the record of the prosodic features of any utterance to be affected by the syntax and content of the utterance as well as by its known position in the discourse.

We may conclude, therefore, that neither the perceptual effectiveness of prosodic end-of-turn cues nor even their existence has been unequivocally established by Duncan's work. We found only two further studies addressing the issue of prosody and turn-taking, both of which preceded Duncan's. The first was a study by Yngve (1970), in which the turn structure of an experimentally elicited two-person conversation was analysed in depth. Although impressionistic, the findings of this study succeeded in ruling out

one plausible hypothesis by establishing that pausing is *not* a turn signal. In other words, it is not the case that speakers simply take over the turn after a sufficient period of silence has elapsed since the last speech from anyone else; rather, they take over when they have received active cues from the previous speaker. The second study, by Meltzer, Morris and Hayes (1971), dealt with only one prosodic dimension: amplitude. Meltzer et al recorded the amplitude fluctuations of individual speakers' voices during sixty two-person problem-solving discussions lasting forty minutes each. Perhaps unsurprisingly, they found that raising one's voice from the normal amplitude baseline correlates well with success at taking over the turn, or keeping it in the face of attempted takeover, and that the absolute difference in amplitude between the two speakers' output efficiently predicts the outcome of an attempted interruption, particularly if simultaneous speech continues beyond a word or two.

3. Methodological Issues in the Study of Turn Signals

These latter two studies raise an interesting question, namely the extent to which one can study conversational structure using non-natural material. Although all the speech Meltzer et al recorded was produced spontaneously, the situation in which it was elicited was experimentally contrived, and designed specifically for collection of the amplitude data they sought. Similarly, Yngve's study used an artificial paradigm in which speakers matched for their conversational ability participated in a conversation designed to be co-operative. Other studies of prosodic factors in conversational interaction, however, have analysed natural speech. Duncan's material was drawn from real-life interviews. French and Local (1982, and this volume) conducted an extensive analysis of natural conversation which produced, inter alia, similar conclusions about the role of amplitude in interruptions to those reached by Meltzer et al. Beattie (1982) and Beattie, Cutler and Pearson (1982) analysed the turn-taking structure of television interviews with politicians; Beattie, Cutler and Pearson analysed prosodic cues in particular. They transcribed a subset of sentence-final phrases "blind", i.e. without reference to the discourse context, and then identified a number of prosodic and vocal quality features which appeared on turn-final and turn-medial utterances respectively. Turn-disputed utterances (i.e. points at which the speaker had been interrupted) could then be analysed in terms of these features, and it could be determined whether they more closely resembled the turn-final or the turn-medial norm.

All such naturalistic studies have one, potentially very serious, drawback: they are based on data from a very limited number of speakers. Duncan's material was produced by three speakers only; Beattie and his colleagues analysed interviews with only two politicians. In contrast, Meltzer et al's experimental study employed 120 speakers; one can be reasonably certain of the generalisability of their amplitude findings. Generalisability cannot, however, be predicated of the naturalistic studies: there is no guarantee that

(a) features characteristic of one speaker's turn-final utterances are also used by other speakers, even other speakers of the same dialect; (b) features which are perceived as effective turn-yielding cues by one listener are effective for others; (c) features which listeners perceive as turn-yielding cues in one speaker's productions are equally effective cues when spoken by others.

The present study forms a first attempt to assess the possibility of using experimental techniques to establish whether perceptually effective prosodic turn signals do indeed exist. Of necessity, the experimental situation was far more constrained than that used by, say, Meltzer et al. In their experiment, only baseline amplitude and excursions from it were at issue; such gross measurements are relatively independent of speech content, so that it was not necessary to constrain the content in any way. Other prosodic characteristics, however - e.g. pitch and timing variation - are more heavily dependent on the speech material in conjunction with which they occur. It is not possible to compare final tone group durations, for instance, when they are realised over different numbers of words. A pitch rise realised on a yes-no question is not necessarily directly comparable with a similar rise realised on a string of words which does not form a question. Thus investigation of such prosodic cues demands careful control of the speech underlying them.

The ideal situation, in fact, would obtain if we had syntactically and semantically identical utterances, produced by the same speaker, which differed only in that one occurred at the end of a conversational turn while the other did not. In the absence of realistically occurring material of this nature, it was therefore decided to approximate it as closely as possible by the simple device of having speakers read aloud short dialogues; the dialogues were written such that the same utterances occurred in either turn-medial or turn-final position in different versions of the texts.

We do not pretend that this experimental design simulates natural conversation. It is, primarily, a device for eliciting the same utterance from the same speaker twice, once in a context in which the speaker is invited to provide turn-final signals and once in a context to which turn-medial signals would be appropriate. By presenting the resulting utterances to listeners, we can determine whether or not listeners fasten upon any particular prosodic features to guide their judgements as to whether a particular utterance is turn-final or turn-medial. In addition, of course, we can determine whether or not speakers do systematically distinguish their turn-final from their turn-medial utterances by prosodic means. Note, however, that this latter issue- the production question - is much less well addressed by the experimental design than is the former - the perception question. The situation in which speakers produce the experimental utterances is far removed from normal spontaneous conversation. The speech is not in the least spontaneous - the task of reading aloud contains at least as large a perceptual as a productive component, and, because the message is given, reading aloud short-circuits the message-

formulation stages of the normal production process. Although professional actors may possibly be able to produce a full range of natural prosodic turn signals when reading a written text aloud, our speakers were untrained in such arts. Even were we to find consistent prosodic differentiation between turn-final and turn-medial utterances in this experiment, therefore, there is no guarantee that such differentiation would reflect the state of affairs in the speakers' normal conversation.

The listeners, on the other hand, at least have a task which approximates the normal case. Though for them too the experimental situation is perhaps somewhat artificial, all they are required to do is to judge utterances as to whether or not they are turn-final. The very premise of this study, as of all other studies of turn signalling, is that such judgements must regularly be made in the course of normal everyday conversation.

4. Description of the Experiment

4.1 Materials

Five dialogues were constructed, each in two versions. An example dialogue is:

Speaker 1:	Foster was pretty upset that you rejected his design - any particular reason?
Speaker 2:	It's simply not good enough, and that's all I have to say on the subject! I don't see why I have to justify my decisions.
Speaker 1:	OK - sorry I asked!

The second version of this dialogue was identical except that Speaker 2's turn read:

Speaker 2:	I don't see why I have to justify my decisions. It's simply not good enough, and that's all I have to say on the subject!

Thus each pair of dialogues provided two sentences which were word-for-word the same, but each was turn-final in one version of the dialogue, turn-medial in the other. The complete set of dialogues is listed in Appendix 1.

4.2 Production Task

Both versions of each of the five dialogues were read onto tape by ten native speakers of British English (six males, four females). For each dialogue, five speakers read one version first while the other five read the other version first. The speakers were instructed to read the dialogues in a natural manner.

Each of the two crucial sentences from each of the ten recordings of each of the two versions of each of the five dialogues was then spliced out of its original context, digitised, and recorded on disc in a computer. The extracts from the example dialogue above, for instance, were "I don't see why I have to justify my decisions" and "that's all I have to say on the subject", each read twice (once turn-finally and once turn-medially) by each of the 10 speakers. There were 200 utterances in all.

4.3 Text Perception Task

Although each dialogue used in the experiment was constructed in such a way that both orderings of the two crucial sentences sounded quite natural, it was nevertheless possible that some individual sentences, out of context, sounded more or less intrinsically turn-final than others. Since in the audio perception tasks the sentences were to be presented out of context, gross differences between the utterances in the "finality" of their texts alone could bias the listeners' judgements. Accordingly we collected "finality" judgements on the sentence texts alone.

The ten crucial sentences were presented in written form to twenty-three native English speakers, none of whom had participated in the production task. They were asked to rate each sentence on a scale from 1 to 5, where 1 represented "definitely still has more to say", 2 "probably still has more to say", 3 "could be going on or could be finished", 4 "probably finished" and 5 "definitely finished". Since the sentences had been chosen to fit equally naturally into turn-final or turn-medial position in context, the ideal result would be a mean rating of 3 for each sentence. In fact, there is one confounding factor which renders this unlikely: although any sentence in the language can be uttered in turn-medial position, i.e. can be followed by some other utterance, not all sentences are suitable for turn-final position. Thus any randomly selected sample of sentences might show a slight bias towards medial ratings; but because all our sentences were chosen so that they could (in our judgement) occur naturally in turn-final position, their ratings might be slightly biased towards the final end of the scale. Thus we predict that the mean finality ratings for all sentence texts will lie between 2 and 4, with the overall mean perhaps slightly above 3.

4.4 Audio Perception Task 1: Isolated Presentation

The 200 utterances were presented singly in random order to twenty subjects, all native British English speakers, who judged for each one whether it sounded turn-medial or turn-final. Ten of the twenty subjects were the speakers who had taken part in the production task. None of the twenty had participated in the text perception task. The subjects were tested individually and heard the utterances over headphones in a soundproof cubicle; they signified their decision by pressing one of two response keys.

This task provides the purest test of whether any general cues to turn-finality (or turn-mediality) are available for listeners to use. Moreover, the fact that speakers in the production task also participate in this task, judging their own as well as others' utterances, provides an extended test of the production question; it may be the case that there are cues which are not easily perceptible to others but which will at least be recognised by the speakers themselves.

4.5 Audio Perception Task 2: Paired Presentation

Twenty-seven subjects, all British English native speakers, none of whom had participated in the production, text perception or first audio perception tasks, heard a tape containing all 200 utterances, paired such that both versions of any one sentence by any one speaker occurred together. In half the pairs the turn-final production occurred first, in the other half the turn-medial. The subjects were tested as a group and heard the tape over loudspeakers in a classroom. They were given a response sheet on which the text of each utterance pair was provided, and recorded their judgements by ticking against each sentence in one of two columns labelled "first" and "second" respectively, to signify which member of the pair they considered to have been the turn-final version.

This test should give more scope than the first audio task for speaker-particular turn signals to become obvious to listeners. Although a given utterance may sound ambiguous in isolation, when it is paired with the alternative version of the same text by the same speaker crucial differences between the two might suffice to enable listeners to make a reliable judgement.

5. *Results*

5.1 Text Judgements

The results of the text perception task are given in Table 1. The overall mean was, as predicted, a little above 3 at 3.43. However, it was not the case that all sentences received mean ratings in the middle range; two were obviously biased towards sounding final. These were "that's all I have to say on the subject", with a mean of 4.83, and "we haven't heard a word from him since", with a mean of 4.30. (When these are removed from the calculation the overall mean lies at 3.15, and the range runs from 2.30 to 3.57.)

That's all I have to say on the subject	4.83
We haven't heard a word from him since	4.30
I really should find out what happened to him	3.57
I don't see why I have to justify my decisions	3.52
- the nicest present I've ever had in all my life	3.48
But she's still there	3.39
I told him to get out and never come back	3.39
You have to take it with a pinch of salt	3.04
He stole all our ideas for the new series	2.48
It was a surprise party	2.30

Table 1 Text perception task: mean finality ratings (1 = most medial, 5 = most final)

The mean number of end judgements elicited by a given sentence in the isolated audio presentation experiment, averaged across speakers and versions, correlated significantly with these text ratings: $r(9) = .824$, $p < .01$. However, when the two high-rated sentences were removed, the correlation was no longer significant: $r(7) = .58$, $p < .10$. Thus there were grounds to believe that subjects' perceptions of these two sentences might have been biased towards turn-final judgements. With the remaining eight sentences, however, we may be confident that no intrinsic bias was confounding the effects of the auditory cues.

5.2 Audio Judgements

As pointed out above, the audio judgement results address two separate questions, which we termed the production question and the perception question. The production question - do speakers consistently produce cues to distinguish turn-final from turn-medial utterances? - can be answered by simple calculation of the correctness scores across speakers and utterances. If listeners are able to categorise utterances correctly with respect to turn position, then we have reason to believe that speakers indeed differentiate between turn-final and turn-medial productions of the same sentences in a consistent way. However, this answer is itself dependent on the perception question - do listeners make consistent use of auditory cues to distinguish turn-final from turn-medial utterances? If they do not, the production question cannot be answered by correctness scores, but must be answered by auditory analysis of the utterances themselves.

The perception question, similarly, must receive a positive answer if correctness scores are high. If not, then two distributions of the results are possible: either all utterances receive about 50% turn-final judgements (i.e. listeners cannot decide about any of them), or some utterances consistently receive more and some fewer turn-final judgements (but these are not

necessarily correct judgements). In the first case, the answer to the perception queston is no, and we must attempt to answer the production question by exhaustive auditory analysis of each utterance. In the second case, we can resort to auditory analysis to answer the perception question; and again, two possible approaches present themselves. On the one hand, if one has a hypothesis about which auditory features will be used as turn signals, one can analyse each utterance for the presence of such features and test the prediction that their presence will be associated with particular categorisations. On the other hand, if one has no a priori expectations of particular auditory features, one can select those utterances with a high proportion of, say, turn-final categorisations, and determine whether they have in common any features which are not shared by utterances without high turn-final ratings.

To begin our anlysis of the audio judgements, we computed the overall correctness scores; these are shown in Table 2. As can be seen, in neither experiment were they very different from chance performance (50%). Removal from the analysis of the two sentences which were intrinsically biased towards "end" judgements did not significantly alter the mean proportion of correct judgements.

	All sentences	Unbiased sentences only
Isolated presentation	51.40	51.44
Paired presentation	53.44	52.60

Table 2 Audio perception tasks: mean percent correct

Interestingly, the ten speakers were no better at judging their own utterances than others were at judging the same utterance, or than they were at judging the others' utterances. No speaker's mean correctness score for his own utterances exceeded 65%, and the mean own-utterance score across the ten speakers was exactly 50%.

There was also no difference between male and female speakers with respect to proportion of correct judgements elicited by their utterances.

Thus the correctness scores force us to resort to auditory analysis to assess the implications of our results. First, however, we must decide whether our auditory analysis should primarily attack the production question (by comparing turn-final with turn-medial productions) or the perception question (by comparing utterances judged as turn-final with those judged to be turn-medial). As we pointed out above, there is reason to believe that the production question might be the less profitable one to investigate, since the nature of the experimental situation did not encourage speakers to indulge in natural speech behaviour. Moreover, the perception judgements lead us to

suspect that listeners' responses were by no means random. We mentioned above that approximately 50% end judgements for all utterances would necessitate an emphatic no to the perception question; but this was definitely not the case. In the isolated presentation experiment the range of end judgements per utterance varied from 0% to 100%. In addition, the mean scores for each utterance pair show a significant, albeit small, positive correlation across the two audio experiments: $r(99) = .221$, $p < .03$. This suggests that the two sets of listeners were making similar decisions about each utterance pair. Finally, the speakers' poor performance at categorising their own utterances also weighs against choosing the production question, since one might expect that if the speakers had in the production task been distinguishing systematically between turn-final and turn-medial productions, they ought to be able in the perception task to detect whatever distinctions they had made. For these reasons we decided to address our further analyses to the perception question.

Since the literature we reviewed above had not provided us with clear hypotheses as to the specific prosodic phenomena involved in turn signalling, we chose not to analyse all utterances and test the correlation between particular prosodic features and particular patterns of listener judgements. Instead, we concentrated upon certain utterances which listeners had clearly perceived to be turn-final or turn-medial. In the isolated audio presentation experiment, 26 of the 200 utterances had been judged turn-medial by 75% or more of the listeners, and 39 had received 75% or more turn-final judgements. Among the latter group, however, we were not surprised to find a total of 24 productions of those two sentences which had proved in the text perception task to be intrinsically biased towards turn-final judgements. In view of the likelihood that categorisations of these utterances were influenced as much by their content as by their prosodic characteristics, we decided to exclude them from this analysis. Thus we had 41 utterances which our listeners had felt to be clearly turn-final or clearly turn-medial.

Before any of the perception experiments had been run, a prosodic transcription of all 200 utterances had been prepared by one of the authors, without knowledge of the context from which each utterance had been taken. The transcriptions of the 41 perceptually unambiguous utterances were now selected and inspected. It was immediately obvious that the utterances which had attracted turn-final judgements had quite different pitch contours from the utterances which had been judged turn-medial. Briefly, turn-final judgements were associated with downstepped contours on the final tone group of the utterance, while turn-medial judgements were associated with final tone groups having upstepped contours. By downstep we mean a tonic syllable starting significantly lower than the previous syllable -Crystal's "drop" and "low drop" (1969, 144-5). Upstep refers to a tonic syllable which starts on a higher pitch than the previous syllable - Crystal's range of "boosters" (1969, 145). Table 3 shows the distribution of upstepped and

downstepped contours across the entire corpus of utterances (excluding those which were textually biased). The distribution is significantly different from chance ($X^2 = 38.3$, $p < .001$).

	75% or more turn-final	Ambiguous utterances	75% or more turn-medial
Upstep	0	38	99
Downstep	100 (N=15)	19 (N=119)	1 (N=26)

Table 3 Percentage of utterances with stepped contours

Figure 1 shows pitch contours and amplitude traces of both versions of the utterance "It was a surprise party" by speaker PC. The top version, with the upstepped contour, received mostly "middle" judgements, while the bottom version, with the downstepped contour, received mostly "end" judgements.

As Table 1 shows, some of the ambiguous utterances also had stepped contours. Inspection of the perception data for these showed that even in this middle range there was a tendency for utterances with downstep to receive more turn-final judgements than utterances without, and for utterances with upstep to receive more turn-medial judgements than those without. We are in no doubt that this feature was used by our listeners as a basis for categorising utterances as turn-final or turn-medial.

No other clear differences between the set of utterances judged turn-final and the set judged turn-medial could be observed in the transcriptions. We also made a range of acoustic measurements on this subset of our data, but on none of them did the two sets of utterances differ significantly, nor did any of the measures correlate with the perception judgements. There was a slight tendency for utterances eliciting a high proportion of turn-final judgements to have a greater overall duration (mean: 2.25 sec) than their pairs which were not considered particularly turn-final (mean: 2.07sec), but this difference also failed to reach our criterion of statistical significance ($F = 4.66$, $p < .07$).

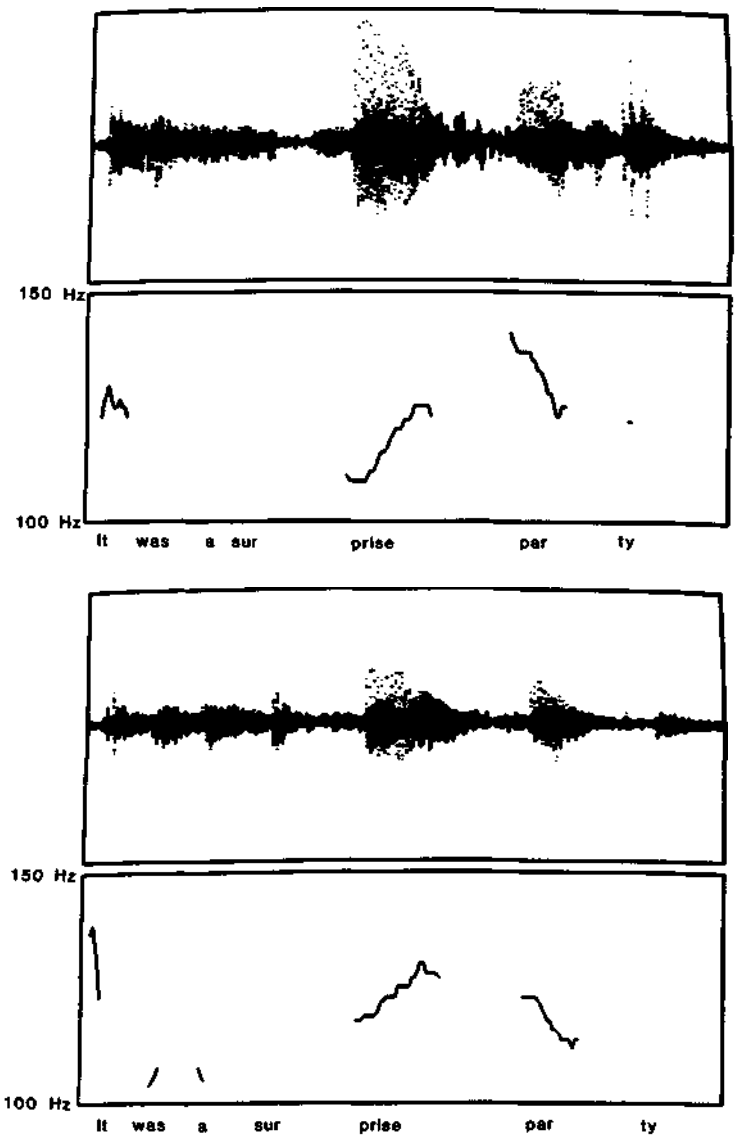

Figure 1 The sentence "It was a surprise party", spoken twice by the same speaker. The top version received mostly "middle" judgements, the bottom version mostly "end" judgements.

6. Conclusion

Our attempt to extend the methodology available for the analysis of turn-taking signals into the laboratory, by using precisely controlled speech output, has met with only qualified success. Our speakers did not differentiate consistently between turn-medial and turn-final utterances, so that we have not been able to give a positive answer to what we termed the production question. However, as we pointed out above, the responsibility for this lies in the shortcomings of our experimental technique, not necessarily in limitations on speakers' use of prosody.

We have, on the other hand, provided clear evidence that the answer to the perception question must be yes. This in itself implies that a positive answer to the production question should be attainable with appropriate methods - if listeners have learned to use cues to turn structure, they surely must have learned by being exposed to cues produced by speakers.

One major cue, we have found, is carried by the fundamental frequency contour of an utterance. The process is more complicated, though, than is suggested by Duncan's (1972) statements that any terminal contour other than a sustained mid-level pitch functions as a turn-yielding signal; our listeners found a downstep in pitch a good turn-yielding cue but a pitch upstep a good turn-holding cue.

This is clearly not the whole story, because many of the utterances which our listeners found ambiguous also had upstepped or downstepped pitch. Other prosodic or vocal quality features which occurred on these utterances may also function as effective turn-holding or turn-yielding cues, and may have cancelled out the stepping contour effects. In natural speech, therefore, there may be quite a range of further turn signals available to speakers and listeners; description of the entire repertoire is a task for future analyses.

However, we are not surprised to have established a clear effect of pitch contour on the perception of turn signals, since there is a body of independent evidence which points to the importance of fundamental frequency contours in discourse structure. Brown, Currie and Kenworthy (1980), for example, found that speakers changing the topic of a conversation signalled this by raising the pitch of their utterance in comparison with their previous pitch level. Exactly the same finding emerged from a study by Menn and Boyce (1982) of parents' conversations with their children. Menn and Boyce also found that a pitch rise (expressed in relation to a speaker's baseline) in comparison with the previous speaker's utterance accompanied any disruption of discourse structure; for example, verification questions,-requiring the conversation to back up temporarily, produced as much pitch raising as did a topic change.

Of course, fundamental frequency is not the only prosodic dimension in which turn signals may manifest themselves. Future studies could well also find turn-taking cues in the amplitude contour of an utterance, since as Meltzer et al (1971) and French and Local (1982) found, amplitude variation plays a large role in determining the outcome of interruption attempts; and Goldberg (1979) also found that speakers changing the topic of discussion tended to do this with an utterance of higher amplitude than their previous utterance. Durational effects specific to the ends of larger discourse units have yet to be established; although the slight tendency in our results for longer utterances to be judged turn-final suggests that here too more sensitive experimentation may be able to establish a perceptual effect.

Methodologically, our experimental paradigm did not prove itself as a useful alternative to natural conversation for studying the production question. However, we did demonstrate that the perception question can be attacked by the rather artificial means of having listeners judge utterances heard in isolation. Accordingly, we recommend the following combination of methodologies for a definitive investigation of prosodic cues to turn-taking: (a) analysis of natural conversation to answer the production question by establishing a repertoire of features characteristic of turn-final and turn-medial utterances respectively; (b) use of speech resynthesis techniques to impose each of these sets of features on otherwise identical utterances, thus creating a range of carefully controlled stimuli which would allow one to ask not only whether a given feature was an effective cue, but which cues were relatively more and which relatively less important. By such means the perception question could be answered in far greater detail than was possible in the present study.

Footnote

This research was supported by the Science and Engineering Research Council and the Social Science Research Council. We are grateful to Stephen Isard andDonia Scott for helpful discussions, and to Peter Clifton for stepping in at the right time.

Appendix 1

Dialogues used in the experiment:

1. *In an advertising agency: an executive and his cantankerous boss*

A: Foster was pretty upset that you rejected his design! Any particular reason?
B: It's simply not good enough, and that's all I have to say on the subject! I don't see why I have to justify my decisions.

OR

I don't see why I have to justify my decisions. It's simply not good enough, and that's all I have to say on the subject!
A: OK, OK, sorry I asked!

2. A shop assistant and one of her regular customers

A: Hello, you're looking particularly happy today!
B: Well, I am happy! It was my birthday yesterday, and I got the nicest present I've ever had in all my life. I got home, and there were about thirty of my friends, all waiting for me - it was a surprise party!

OR

Well, I am happy! It was my birthday yesterday, and when I got home, there were about thirty of my friends, all waiting for me - it was a surprise party! I think it was the nicest present I've ever had in all my life.
A: Isn't that nice! No wonder you look pleased!

3. *Two neighbours on campus*

A: I haven't seen that friend of yours lately - Roger, was that his name? He used to be round here all the time.
B: You know, that's a funny thing. I was thinking myself, only the other day, I really should find out what happened to him. You see, he went off to Africa - it was supposed to be just for a holiday - and we haven't heard a word from him since.

OR

You know, that's a funny thing. He went off to Africa - it was supposed to be just for a holiday - and we haven't heard a word from him since. I was thinking myself, only the other day, I really should find out what happened to him.

A: That's weird. I hope he's all right.

4. *Two colleagues in a TV studio*

A: I hear you and Joe had a bit of a row.

B: You're damn right we did! Do you know what the bastard did? He stole all our ideas for the new series! I told him to get out and never come back!

OR

You're damn right we did! I told him to get out and never come back! Do you know what the bastard did? He stole all our ideas for the new series!

A: Hey, you know this isn't the first time that's happened with him. Why do you think he lost his job in London?

5. *Conversation between two friends*

A: Hey, I saw Carol at the weekend.

B: Oh yeah?

A: She was really pissed off with living in that flat. She reckons she's moving out, and if Chris wants to stay there, that's up to him.

B: Well, so she says, but I think you have to take it with a pinch of salt. I don't know how many times she's sworn she was moving out - but she's still there.

OR

Well, so she says. I don't know how many times she's sworn she was moving out - but she's still there. I think you have to take it with a pinch of salt.

A: Well, she seemed pretty determined to me.