Published paper

Kurtic, E., Brown, G. J. and Wells, B. (2009) *Fundamental frequency height as a resource for the management of overlap in talk-in-interaction.* In: Where Prosody Meets Pragmatics. Studies in Pragmatics, 8 . Emerald Group Publishing Limited , Bingley, UK ISBN 978-1-84950-631-1

Where Prosody Meets Pragmatics
Edited by Dagmar Barth-Weingarten, Nicole Dehé and Anne Wichmann

8

Fundamental Frequency Height as a Resource for the Management of Overlap in Talk-in-Interaction

Emina Kurtic, Guy J. Brown and Bill Wells

Abstract

Overlapping talk is common in talk-in-interaction. Much of the previous research on this topic agrees that speaker overlaps can be either turn competitive or noncompetitive. An investigation of the differences in prosodic design between these two classes of overlaps can offer insight into how speakers use and orient to prosody as a resource for turn competition.

In this paper, we investigate the role of fundamental frequency (F_0) as a resource for turn competition in overlapping speech. Our methodological approach combines detailed conversation analysis of overlap instances with acoustic measurements of F_0 in the overlapping sequence and in its local context. The analyses are based on a collection of overlap instances drawn from the ICSI Meeting corpus. We found that overlappers mark an overlapping incoming as competitive by raising F_0 above their norm for turn beginnings, and retaining this higher F_0 until the point of overlap resolution. Overlappees may respond to these competitive incomings by returning competition, in which case they raise their F_0 too. Our results thus provide instrumental support for earlier claims made on impressionistic evidence, namely that participants in talk-in-interaction systematically manipulate F_0 height when competing for the turn.

1. Introduction

Overlapping speech is a common phenomenon in naturally occurring conversations. Given that for the most part conversations proceed smoothly, without overlaps, the occurrence of overlap in conversation and its management by conversational participants require explanation.

According to the influential model of turn-taking by Sacks *et al.* (1974) conversation participants aim to minimise gaps and overlaps in conversations. Overlapping speech instances are described as 'common, but brief', and the briefness is explained by the fact that overlaps are most often placed at possible turn-ends, around a so-called *transition relevance place* (TRP) where the current speaker should terminate his or her turn (Sacks *et al.*, 1974). According to this model, overlaps commonly occur as a result of self-selection and projectability of turn-ends. Self-selection occurs in cases where the current speaker does not select the next speaker, so that one or more participants may self-select, potentially giving rise to a simultaneous start. Alternatively, a participant may self-select as next speaker before actual completion of the turn, but at the point where such completion is projected. Jefferson's work on precision-timing (Jefferson, 1973) provides evidence that conversation participants are monitoring the progress of the ongoing turn and are able to time their turn beginning precisely at the current speaker's turn end, and thus avoid overlap (or a gap). If overlap occurs before this point, but still in the space where this point is projected, it is likely to be the result of mistiming. Thus the model of Sacks *et al.* (1974) accounts for the occurrence of overlap within the TRP space: the overlap is explained as resulting from turn-taking principles.

Many subsequent studies on overlap in conversation contrasted such overlap instances that seem to result from regular turn-taking mechanisms with those in which participants compete for the turn in progress (e.g. French and Local, 1983; Jefferson, 1983; Couper-Kuhlen, 1993; Schegloff, 2000; Wells and Corrin, 2004). In French and Local (1983), turn competitive overlaps are defined as those instances in which the incomer is heard as 'wanting the floor to him/herself not when the current speaker has finished but now at *this* point in conversation'. Schegloff (2000) characterises these overlaps as those instances in which the conduct of participants demonstrates that they treat the in-overlap speech as problematic and in need of resolution. Turn competition does not have to be confined to the incoming speaker: as Schegloff (2001) puts it – where there is the attempt to 'drive the prior speaker out', it can be the aim of either party.

However, not all overlap is turn competitive. In addition to overlaps at the TRP, described above, conversations contain a large number of noncompetitive overlaps that have different conversational functions. One common class of noncompetitive overlap are the so-called continuers (Schegloff, 1982) or backchannels (Yngve, 1970) which are commonly used by overlappers to mark the recipiency of the ongoing turn and confirm the current speaker's right to an extended turn. Schegloff (2000) lists three further types of overlaps: *terminal overlaps* (i.e. overlaps at the TRP), *collaborative completions* and *choral productions* in which generally no evidence of turn competition can be found.

Given that these two broad types of overlaps, competitive and noncompetitive, are well attested in the literature, an obvious next task is to investigate what linguistic resources are employed by the conversationalists in order to display an overlap as turn competitive or as noncompetitive. This question has been addressed by several previous studies in conversation analysis (CA).

Jefferson (1983) investigated the precise placement of overlap onsets and found that they may occur systematically at *any* place in the ongoing turn. According to her, the positioning of the overlap onset is related to the competitiveness of the overlap. She offers a preliminary categorisation of overlap onsets according to their position relative to the TRP into *transitional*, *progressional* and *recognitional* onsets. Transitional onsets are located at the TRP, whereas progressional onsets start at the silence after an uncompleted utterance. In Jefferson's terminology (Jefferson, 1983, p. 28) these overlaps are called 'byproduct overlaps' as they are a byproduct of routine turn-taking practices (as described by Sacks *et al.*, 1974). Recognitional onsets are located at points where the incoming speaker has gained sufficient understanding of the current speaker's turn. These onsets result in so-called 'first-order overlaps of varying degrees of turn incursion' (Jefferson, 1983, p. 28). Jefferson's differentiation between 'byproduct' and 'first-order' overlaps thus corresponds to noncompetitive and turn competitive overlaps respectively.

According to French and Local (1983), the placement of overlap onset within the current speaker's talk is not relevant for characterisation of overlap as turn competitive or not. They also argue against the overlap's lexical design and the pragmatic function (i.e. being an agreement or disagreement) as being robust features for discrimination between competitive and noncompetitive overlaps. According to their analysis, it is the combination of raised pitch and volume (abbreviated to $<\mathrm{h}+\mathrm{f}>$) that fulfills this function. French and Local (1983) offer evidence that $<\mathrm{h}+\mathrm{f}>$ is utilised by the overlapping speakers (henceforth, overlappers) to compete for the turn, and is also treated as competitive by the turn-holders (henceforth, overlappees).

Pitch and volume have also been reported in connection with overlap management by Schegloff (2000) and Shriberg *et al.* (2001a). Schegloff (2000) regards increases in pitch or volume as turn competitive 'hitches' that occur in competitive overlaps. In a quantitative study of overlaps on a large corpus, Shriberg *et al.* (2001a) report higher fundamental frequency (F_0) and energy at the onsets of turns in overlap, compared to the onsets of turns from silence (i.e. not in overlap). However, as their study did not differentiate between competitive and noncompetitive overlaps, the conversational function of these prosodic resources remains unclear.

The relationship between positioning of overlap onset and prosodic design of the incoming is investigated by Wells and Macfarlane (1998). Synthesising the analyses by French and Local (1983) and Jefferson (1983), they claim that $<\mathrm{h}+\mathrm{f}>$ is the major indicator of turn competitiveness, and that incomings having this prosodic design are positioned before the last major accented syllable in the current speaker's turn (Wells and Macfarlane, 1998: 272).

Positioning before the major accented syllable alone does not indicate competition, as shown by overlaps starting at the points where the current speaker is disfluent. These incomings can be placed before the major accented syllable, but do not seem to display $< h + f >$, in which case they are not treated as turn competitive in spite of their placement.

In the current study, we focus on F_0 as a resource for turn competition. Specifically, our analysis aims to answer the following questions:

- Do overlappers use higher F_0 for competitive overlaps than for noncompetitive overlaps?
- Do overlappees modify F_0 during competitive overlaps as opposed to noncompetitive overlaps?
- Are the F_0 modifications made by overlappees dependent on their interactional response to a competitive incoming?

In this respect we build upon those studies reported above, that use a combination of CA and phonetic analysis. Like those studies, we use audio recordings of naturally occurring spontaneous spoken interaction as a basis for interactional and phonetic analysis. However, our study departs from them with respect to the method of phonetic analysis. Previous studies relied principally on impressionistic listening rather than acoustic analysis, not least because of the technical challenges of analysing instrumentally the simultaneous speech signals from two or more speakers. In the current study, we address this issue in two ways. Firstly, we use a corpus of audio recordings in which the individual conversational participants are recorded on separate audio channels (see Section 2 below). Secondly, we take advantage of audio signal processing algorithms that are able to reliably track the fundamental frequency of speech in the presence of background noise or another voice (de Cheveigné, 2006). By placing a greater emphasis on instrumental measurement of spoken interaction, we aim for a more reliable and objective characterisation of the phonetic features that are implicated in the management of overlapping talk.

In the present paper the investigation is restricted to measurement of F_0, which is generally taken to be the main (though not only) correlate of perceived pitch (Moore, 2003). We thus focus on just one of the various phonetic parameters that have been hypothesised to play a role in the management of overlapping talk, the others being loudness and tempo. One reason for this decision is that in our collection of overlaps from the ICSI corpus, F_0 is relatively straightforward to measure, compared for example to intensity, which is the main acoustic correlate of perceived loudness. This is because when speech overlaps, it is more difficult to apportion the energy in the mixture to each speaker than it is to track two overlapping F_0s. In addition, F_0 is not affected by variations in microphone placement, whereas intensity is. Although we chose our data to minimise the possibility of variations in microphone placement (by selecting speakers who used headset rather than lapel microphones), variability of sound level due to microphone placement is still likely to be an issue.

2. Method

2.1. Data

Our analyses are based on the ICSI Meeting corpus (Janin *et al.*, 2003), a corpus of spontaneous multi-party research meetings. ICSI meetings are spontaneous conversations in that they are not set up specifically for recording purposes, but rather are regular meetings of a research group. A subset of meetings was selected to include the five (three male and two female) native speakers of American English who were present at most of the meetings that make up the corpus. The corpus contains both meetings directed by one person and those where less constrained exchange between participants is taking place. The present analysis is based on overlaps drawn from two Meeting Recorder meetings (Bmr008 and Bmr016). These are meetings of a less constrained type in which the creation of the ICSI Meeting corpus itself is discussed. These two meetings have six and seven participants respectively, including the selected five speakers.

In these two meetings the total number of overlaps involving only the selected speakers is 860. From these we exclude all overlaps that involve more than two people speaking at the same time (141 overlaps) and analyse only two-speaker overlaps in the interests of simplifying the analysis and for comparability with the results in French and Local (1983). Furthermore, in this study we are, like French and Local, only interested in overlaps ‘in which one speaker comes in clearly prior to the completion of another’s turn’ (French and Local, 1983). For this reason, we also exclude all overlaps that are placed around what conversation participants are likely to interpret as a potential TRP (324 overlaps). These include incomings placed at points of syntactic completion, simultaneous starts, terminal overlaps that begin at the last word of a turn, and so-called ‘blind spot’ overlaps (Jefferson, 1986). We also exclude backchannel continuers (237 overlaps), collaborative completions (12 overlaps) and choral productions (32 overlaps) as these noncompetitive overlap classes have well defined conversational functions, and as such deserve a separate analysis. This leaves us with a set of 114 overlap instances. This substantial reduction in the number of two-speaker overlap instances shows that most two-speaker overlaps either belong to one of the noncompetitive categories (mostly continuers) or are placed around a TRP.

2.2. Overlap identification

Overlap instances were detected automatically using start and end time information for each word that was provided with the corpus. This information was obtained from a forced alignment between the word-level transcriptions of the meetings and the corresponding speech signals using an automatic speech recogniser. The main unit of data segmentation is the turn, as defined in the process of corpus transcription (Edwards, 2004). Each turn is associated with a start and end time that makes it possible to align it with the speech recording. Overlap detection was achieved as follows. First, all turns containing overlaps were identified based on overlapping start and end times at the turn level. Subsequently, word-level forced alignments of the corpus were used to identify which words overlap within a turn. The entire overlap region was delimited by the start time of the first overlapping word and end time of the last overlapping word.

2.3. Interactional analysis: Competitiveness classification

Sequences of overlapping talk were analysed and categorised as competitive or noncompetitive. An overlap is regarded as competitive if it can be shown that conversation participants regard it as such. In order to do so, detailed CA is needed for each instance of overlapping talk. The segment of overlapping speech below (from ICSI Meeting Bmr007) shows an instance of a competitive overlap. According to CA transcription conventions (Jefferson, 2003), the overlaps are indicated by square brackets ([]). '(.)' indicates a pause of duration less than 0.2 sec and '.hhh' marks inbreath.

```
Bmr007_109

    1  m1:  Well but see I find it [interesting]
    2  f2:  [So:]
    3  m1:  even if it wasn't any more (0.2)
    4       because (.) since we were dealing with this full duplex
            sort of thing in Switchboard where it was just all
            separated out
    5       .hhh
    6  f2:  Mm-hmm?
    7  m1:  we just everything was just nice so the (.) so the issue
            is in (.) in a situation
    8       (0.4)
    9       [where th that's ]
→  10  f2:  .hhh [Well it's not really] (.) nice it depends what
            you're doing
   11       So if you were actually
   12       .hhh (0.4) having (0.3) uh (0.5)
   13       Depends what you're doing
   14       if (1.15)
   15       Right now we're do we have individual mikes on the people
            in this meeting
   16       (0.3)
   17  m1:  Mm-[hmm]?
   18  f2:  [So ] the question i:s, >*you know*<
```

```
19       (.)
20       are there really more overlaps happening. hhh
21       (0.9)
22       than there would be in a two-person (0.2)*[party]*.
23  m1:  .hhh [Let ]
24  f2:  And [and there well may be, *but* ]
25  m1:  [let m let me rephrase what I'm saying]
26       cuz I don't think I'm getting it across.
27       What
28       What I what
29       (0.5)
30       I shouldn't use words like ''nice'' because maybe that's
         too i too imprecise.
```

Speaker f2 starts her turn in line 10 at the point in speaker m1's turn that is not a point of syntactic completion. Even though f2 starts the overlap during the final part of m1's turn (*so that-the issue is*...), she chooses to address a preceding part of it (*everything was just nice*) thus attempting to bring the topic back to *nice* and prevent m1 from continuing towards the turn completion. f2's and m1's *nice* do not to refer to the same event as indicated by m1's later request for correction from line 24 onwards. This suggests that the adjective *nice* is selected by f2 as a suitable linguistic focus for an incursion into m1's turn. m1 abandons his turn whereupon f2 secures the floor for an extended turn (ll. 10–16). Despite her many disfluencies and long pauses, no other participants attempt to take over from f2 until m1 claims a turn again in line 23. These positional, syntactic and pragmatic criteria offer evidence of f2's turn competitive behaviour, so this overlap is classified as competitive. In this way, instances of competitive and noncompetitive overlap can be identified independently of their prosodic design.

By analysing the conversation sequences in which they occur, two annotators classified into competitiveness categories a total of 419 overlaps, including the 114 targeted in this study. One annotator had previous training in CA, but no previous experience of overlap classification. This annotator was given the definition of overlap competitiveness along with examples of competitive and noncompetitive overlap. The other annotator was new to CA. She was given the same description as the first annotator and was additionally trained by discussing 100 overlap instances with the first author. The two annotators then classified the overlap set independently and reached an agreement of Cohen's kappa: $\kappa = 0.67$.

This agreement is considered acceptable in general terms (Carletta, 1996) and is within the range of what can be expected for similar dialogue act classification tasks. Although there are no directly comparable competitiveness classification studies, inter-rater agreement is available for several dialogue act classification tasks that were conducted for computational studies of dialogue. For example, the classification of turns as agreement, disagreement, backchannel and other in

Galley *et al.* (2004) results in an agreement of $\kappa = 0.63$. Di Eugenio *et al.* (2000) assess inter-rater agreement on various dialogue annotation tasks and report agreement between $\kappa = 0.54$ (for [accept, reject, hold, $\varnothing$] classification) and $\kappa = 0.79$ (for [answer, $\varnothing$] classification). The closest task to competitiveness classification is the identification of 'floor-grabbers' in the dialogue act annotation of the ICSI Meeting corpus (Shriberg *et al.*, 2004). However, the inter-rater agreement for this annotation task is only reported for groups (maps) of dialogue act categories and ranges from $\kappa = 0.76$ to $\kappa = 0.80$ depending on the number of categories included in a map. Therefore, we do not have inter-rater agreement for 'floor-grabber' alone and cannot directly compare the level of agreement.

Of the 114 overlaps, 80 were judged to be competitive and 34 noncompetitive. The distribution of competitive and noncompetitive overlaps in this final data set reflects the tendency found in the data as a whole, namely that competitive overlaps are placed prior to a TRP more often than noncompetitive ones.

2.4. Acoustic analysis

2.4.1. Extraction of F_0 contours. The ICSI corpus contains an audio channel for each talker, which was used for F_0 analysis. However, such channels are not free from crosstalk from other participants (cf. Wrigley *et al.*, 2005) or from interference associated with other non-speech noises in the room (e.g. rustling paper, closing doors etc.). For prosodic analyses we use speech data recorded on close talking headset microphones to minimise the contamination of the sound by crosstalk and non-speech sounds. Nevertheless, we found that an appreciable amount of crosstalk could still occur during regions of overlapping speech. This needed to be addressed in order to obtain reliable F_0 estimates.

For F_0 extraction, we use the YIN pitch determination algorithm (de Cheveigne and Kawahara, 2002). Standard pitch tracking algorithms such as YIN are expected to track the most prominent F_0, which should correspond to the desired talker (since the level of the talker's speech on their own audio channel was usually substantially higher than the level of the crosstalk). However, in practice we found that F_0 contours obtained using YIN were unreliable during regions of overlapping speech.

Accordingly, F_0 contours were derived semi-automatically in a two-stage process. First, a rough estimate of the fundamental period at each time frame was made by drawing a contour on a visual representation of the speech periodicity. This was based on the 'cumulative mean normalised difference function' $d'(\tau)$ proposed by de Cheveigne and Kawahara (2002), given by

$$d'(\tau) = \begin{cases} 1 & \text{if } \tau = 0 \\ \dfrac{d(\tau)}{[(1/\tau)\sum_{j=1}^{\tau} d(j)]} & \text{otherwise} \end{cases}$$

which is derived from a difference function

$$d(\tau) = \sum_{j=1}^{W} (x_j - x_{j+\tau})^2$$

Here, $x(t)$ is the speech signal and τ a time lag, which was varied within the range of plausible pitch periods (up to a maximum of 20 msec, corresponding to a lower bound on the F_0 of 50 Hz). W is the window length (25 msec) and the index j counts time in steps of the sample period. The first major dip in $d'(\tau)$ is usually a very good indicator of the pitch period.

Subsequently, the rough estimate of the fundamental period was refined by searching $d'(\tau)$ for the local minimum nearest to the estimated period at each time frame, and fitting a quadratic around the minimum in order to get an accurate estimate of the fundamental period. The pitch tracking application also allowed the F_0 contour to be heard as a pure tone whose frequency followed the F_0, which provided an audible check that the F_0 of the correct speaker was being tracked in cases where there was substantial crosstalk.

2.4.2. Unit of measurement. The time window over which F_0 is measured is an important consideration in our study and is closely related to the question of what kind of speech unit underlies the mechanisms of turn competition. French and Local (1983) refer to the 'foot' as being the major unit that participants can design prosodically as competitive or noncompetitive. The foot is an interval between two prominent syllables of speech; hence if the foot was adopted as the unit of measurement in our study it would be necessary to have a computational means of detecting 'prominence' in the speech signal. Detecting prominence automatically, particularly in spontaneous multi-party speech, is still a challenging research problem (e.g. Tamburini and Caini, 2005; Wang and Narayanan, 2006) although the composite measure used by Calhoun (this volume) offers a possible solution. Therefore we adopted the approach used in a number of quantitatively oriented studies of overlapping speech and prosody of discourse (Koiso *et al.*, 1998; Shriberg *et al.*, 2001b; Caspers, 2003) that segment the data into units delimited by short pauses. After Koiso *et al.* (1998) we call these units interpausal units (IPU), and define an IPU as a stretch of speech between pauses of at least 0.1 sec. The main reason for selecting 100 msec as the pause length to define units of analysis is that it has been claimed by researchers in CA, on the basis of close analysis of naturally occurring talk-in-interaction, that the shortest pauses that speakers orient to as interactionally relevant are around a tenth of a second (cf. Couper-Kuhlen (1993: 122–123) for summary). However, as Couper-Kuhlen points out, detecting pauses in speech is not a straightforward matter, so there is inevitably an element of arbitrariness in selecting a particular length of pause.

Figure 8.1 illustrates the segmentation of an overlap sequence into IPUs. Three IPUs in the local overlap context are relevant for the analyses: IPU_b (the IPU immediately preceding the

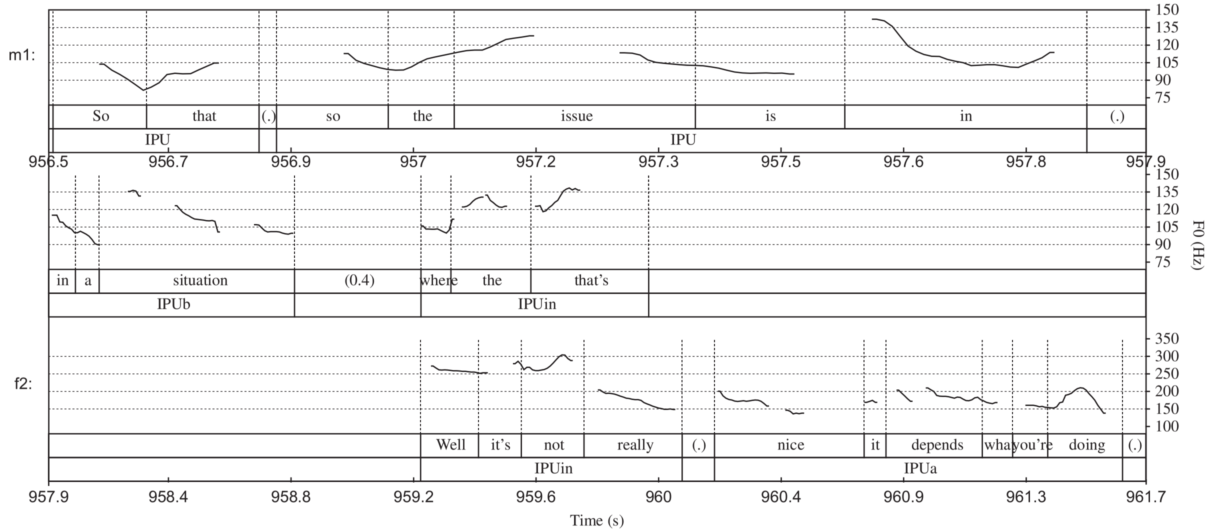

Figure 8.1. An example segmentation into interpausal units (IPUs) for the competitive overlap instance shown in the extract above (ll. 7–10) also showing F_0 tracks for this overlap segment. In this particular example there is one IPU (IPU_{in}) between the start and the end of overlap. In the IPU before the overlap (IPU_b) the overlapper is silent, and in the IPU after the overlap (IPU_a) the overlapper continues (i.e. wins the floor).

overlap onset point), IPU_{in} (the first IPU of the overlap) and IPU_a (the IPU following the point of overlap resolution). The F_0 is extracted at 10 msec intervals within each IPU.

3. Results

Applying the methodology outlined above we investigate how overlappers use fundamental frequency height to compete for the turn, and how overlappees respond to competitive and noncompetitive incomings respectively by varying F_0 height. The results of the five speakers are pooled in order to provide sufficient power for statistical analysis, as a step towards characterising the use of F_0 variation in overlapping talk in Mainstream American English. We also present results for each of the five-targeted speakers separately, to give an indication of individual differences in turn competition behaviours.

3.1. F_0 height of overlapping incomings

The first question we address is whether F_0 height is used by incoming speakers (overlappers) to compete for the turn. If overlappers use F_0 to compete for the turn, the F_0 design of competitive incomings should be different from that of regular, non-overlapping turn starts in a smooth turn exchange. We further expect the F_0 design of competitive incomings relative to this norm for turn beginnings to be different from that of noncompetitive incomings.

Another prediction is that if F_0 height is used as a resource for turn competition, its use will be limited to the amount of time during which competition is taking place (French and Local, 1983). After resolution of competition that coincides with resolution of overlap itself, we expect to find a change in F_0 again.

To test these predictions, we measure the mean F_0 of the first in-overlap IPU in the audio recording for each overlap instance and compare it to two reference values: the mean F_0 of turn starts in non-overlapping talk and the mean F_0 of the IPU immediately following the overlap. All F_0 values are presented in semitones (ST) relative to 16.35 Hz (Baken and Orlikoff, 2000).

Table 8.1 shows the results of an independent *t*-test for the significance of difference in F_0, in ST, between overlap onset and the norm for non-overlapping turn beginnings in competitive and noncompetitive overlaps.

The results show that on average the F_0 is higher at the overlap onset than at the beginning of a turn after a smooth transition. This difference is significant for competitive incomings, but not for noncompetitive incomings. This suggests that speakers use F_0 relatively higher than their norm for turn beginnings to start competitive incomings, but not to start noncompetitive incomings. However, the sample size is a good deal smaller in the case of noncompetitive incomings, and

Table 8.1. Overlappers: Results of an independent *t*-test for the significance of difference between F0 at the overlap onset and the norm for non-overlapping turn beginnings in competitive and noncompetitive incomings

	Context	**N**	**Mean F_0 (ST)**	**SE**	**Significance of difference**
Competitive	Overlap onset (IPU_{in})	80	39.4660	0.67427	$t(197) = 3.470$, $p < 0.01$
	Norm for turn begins	119	36.6034	0.50213	
Non-competitive	Overlap onset (IPU_{in})	34	38.3841	1.13885	$t(151) = 1.592$, $p = 0.159$
	Norm for turn begins	119	36.6034	0.50213	

there is a higher standard deviation. It therefore cannot be ruled out that with a larger sample, the F_0 of noncompetitive incomings would also prove to be significantly higher than the norm for turn beginnings.

Figure 8.2 shows the pattern of results for individual speakers. Zero represents the normalised mean F_0 for the start of non-overlapping turns (i.e. ‘in the clear’). The bars show the mean variation from that norm, in ST, for each speaker. *n* refers to the total instances of overlap for that speaker. The figure shows that for four out of five speakers, both competitive and noncompetitive overlaps are realised with a higher F_0 than turn initiations in the clear (i.e. not in overlap). In all four cases, the F_0 of competitive overlaps is higher than that of noncompetitive overlaps. In the case of speaker m2, similarly, the F_0 of the competitive overlaps is higher than noncompetitive overlaps. However, for this speaker, both types of overlap are realised with an average F_0 that is lower than that of turn initiations in the clear.

Next we turn to the question of whether there is a difference in the mean F_0 between the overlap onset and the first IPU following the overlap.[1] The existence of the overlappers’ post-overlap IPU means that the overlappee does not continue past the overlap resolution point in this

[1] Since the large majority of overlap instances in this set contain just one in-overlap IPU, in most cases the ‘overlap onset IPU’ is coextensive with the entire overlap.

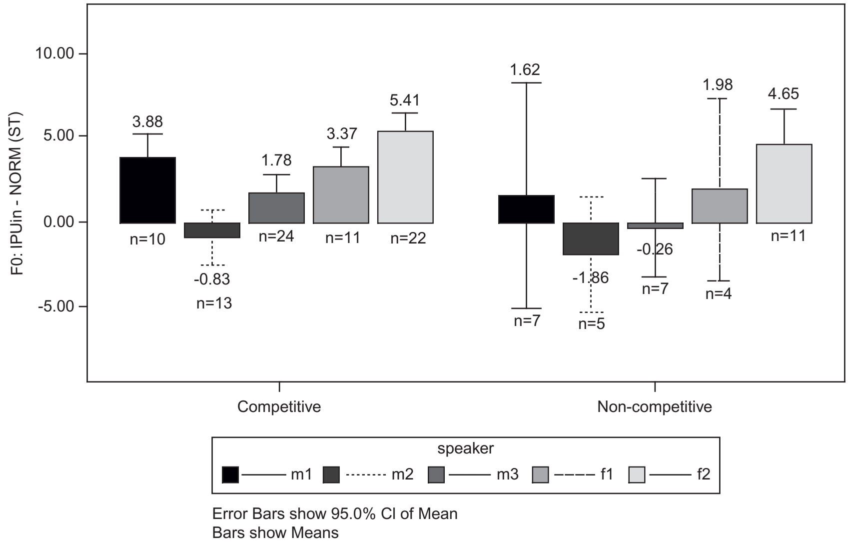

Figure 8.2. Mean F_0 difference (in semitones) between the first IPU of the overlap and the norm for turn beginnings for each speaker in competitive and noncompetitive incomings.

particular turn, either because he/she loses turn competition or because he/she yields the turn to the incomer without competition. This is the case in 65 out of 80 competitive overlaps and 21 out of 34 noncompetitive overlaps. Table 8.2 shows the results of a dependent *t*-test that was used to assess the significance of the differences between overlap onset IPU and the post-overlap IPU.

In competitive overlaps the mean F_0 of the overlap onset is significantly higher than the mean F_0 of the first post-overlap IPU. It could be hypothesised that the fall in F_0 from the overlap onset IPU to the first post-overlap IPU is attributable to the well-described natural declination of F_0 across the turn. However, this is contradicted by the fact that in noncompetitive overlaps we find no significant drop in F_0 from the overlap onset IPU to the first post-overlap IPU. Moreover, there is no significant difference in F_0 height of the post-overlap IPU between competitive and noncompetitive overlaps ($t(91) = 0.559$, $p = 0.578$), which means that F_0 falls to a similar level after turn competition as it does when no turn competition takes place. This suggests that incomers' speech is kept higher in F_0 for the duration of the overlap and is lowered to a level that seems to be the norm for post-overlap F_0 level when the competition is resolved.

Figure 8.3 shows that this general pattern is reflected in each of the individual speakers: for each speaker, the difference in mean F_0 between the portion in overlap and the following, non-overlapped talk is greater when the overlap is competitive.

Table 8.2. Overlappers: Results of dependent *t*-test for the significance of difference between F0 (in semitones) at the overlap onset and the first post overlap IPU in competitive and noncompetitive incomings

	Context	N	Mean F_0 (ST)	SE	Significance of difference
Competitive	Overlap onset (IPU_{in})	65	38.7912	0.73974	$t(64) = 7.671$, $p < 0.001$
	Post overlap (IPU_a)	65	36.4406	0.74381	
Non-competitive	Overlap onset (IPU_{in})	21	36.8607	1.29270	$t(20) = 1.131$, $p = 0.271$
	Post overlap (IPU_a)	21	35.9983	1.29685	

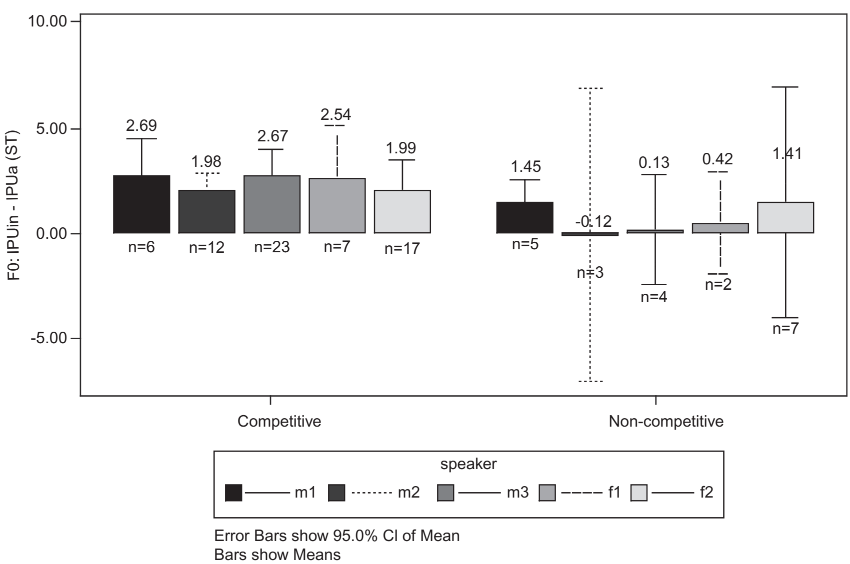

Figure 8.3. Overlappers: Mean F_0 difference (in semitones) between the first IPU of the overlap (IPU_{in}) and the IPU following the overlap (IPU_a) for each speaker in competitive and noncompetitive incomings. Zero represents the mean F_0 of the IPU_a.

3.2. F_0 height of overlapped turns

The results for overlappers suggest that raised F_0 is a systematically deployed resource for turn competition. We now consider whether overlappees alter their F_0 in response to overlapping incomings.

This question is addressed by comparing the mean F_0 of the first IPU in overlap to the F_0 of the IPU immediately preceding the overlap. In the case of noncompetitive overlaps, there is no reason to predict an alteration in F_0, since the overlappee's claim to the turn is not under threat. In the case of competitive overlaps, at least two possible interactional responses of the overlappee can be envisaged: either he yields the floor to the incomer, or he returns competition (French and Local, 1983). Overlappee responses to competitive incomings were therefore subdivided into two categories: turn-yielding versus returning competition. This classification was based on the original annotation described above. The set of overlaps in which overlappee returns competition contains both cases in which he/she is successful in competing for the turn and continues past overlap resolution point, and those in which he/she loses the competition and gives up the turn. Overlaps in which overlappee yields the turn are identified as such by an absence of sequential evidence of competitive behaviour.

Table 8.3 gives the results of a dependent *t*-test for both subcategories of competitive overlaps and noncompetitive overlaps.

The results show that on average overlappees significantly lower their F_0 compared to the IPU preceding the overlap, both in response to noncompetitive incomings and when not returning competition to competitive incomings. Where overlappees return competition, the difference in F_0 means is not significant.

Figure 8.4 gives single speakers' values for the overlappees' response to overlap types presented in Table 8.3. The figure shows that all speakers lower F_0 below the values of pre-overlap speech when they are overlapped noncompetitively or when they do not return competition. However, speakers differ in the way they design their competitive response to competitive incomings. Three out of five speakers (m1, m3 and f2) lower F_0 in a similar way as they do upon noncompetitive incomings or when not returning competition. The remaining two speakers raise F_0 above the level of preceding IPU thus marking the difference between their noncompetitive and competitive responses by this F_0 modification.

4. Discussion

This work investigated how conversation participants' use and orient to F_0 height in order to manage turn competition in overlapping speech. For this purpose we have investigated how both

Table 8.3. Overlappees: Results of dependent *t*-test for the significance of difference between F0 at the overlap onset and the IPU preceding the overlap upon competitive and noncompetitive incomings

		Context	N	Mean F_0 (ST)	SE	Significance of difference
Competitive	Overlappee returns competition	Pre overlap (IPU_b)	39	38.8986	0.90878	$t(38) = 0.919$, $p = 0.364$
		IPU upon overlap onset (IPU_{in})	39	38.4463	0.95413	
	Overlappee yields the turn	Pre overlap (IPU_b)	41	35.9536	0.95207	$t(40) = 2.795$, $p < 0.01$
		IPU upon overlap onset (IPU_{in})	41	34.8158	0.85070	
Noncompetitive		Pre overlap (IPU_b)	34	38.4014	1.34231	$t(33) = 3.584$, $p < 0.01$
		IPU upon overlap onset (IPU_{in})	34	36.4129	1.13352	

overlappers and overlappees deploy F_0 height in overlap, compared to their regular turn starts and also compared to the local context within the turn in which overlap occurs.

- Do overlappers use higher F_0 for competitive overlaps than for noncompetitive overlaps?

We first considered this question with reference to the *onset* of the overlap. For our speakers as a group, overlapping incomings are routinely marked as competitive by an increase in F_0 above the norm for turn beginnings. With regard to individual speakers, this was true of four of the five

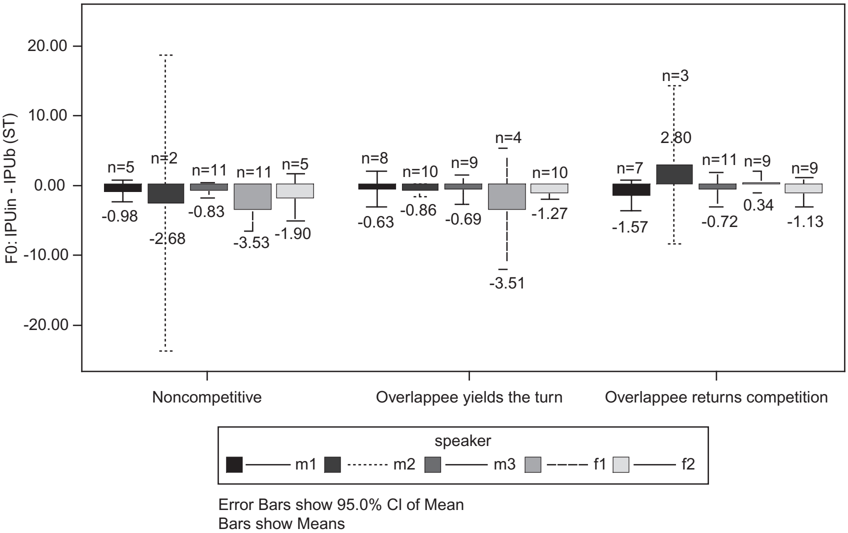

Figure 8.4. Overlappees: Mean F_0 difference (in semitones) between the first IPU of the overlap (IPU_{in}) and the IPU preceding the overlap (IPU_b) for each overlappee when she/he returns the competition and yields the turn without competition. Zero represents the mean F_0 of the IPU_b.

speakers analysed. The fifth speaker did not routinely raise F_0 above the norm from competitive incomings; however, for him F_0 was nevertheless consistently higher for competitive than for noncompetitive overlaps. We can therefore conclude that overlappers do use relatively higher F_0 to start competitive overlaps, compared to noncompetitive overlaps.

We next considered whether incomers maintain higher F_0 through the course of a competitive overlap, until the point of overlap resolution. We found that incomers' speech is kept higher in F_0 for the duration of the overlap and is lowered to a level that seems to be the norm for post-overlap F_0 level when the competition is resolved. This was the case for each individual speaker and for the group as a whole.

It can therefore be concluded that, at least for these speakers of Mainstream American English, relatively high F_0 is a routine feature of the design of turn competitive incomings. While there are considerable differences among the five speakers in the *amount* of variation in F_0 height that they deploy, all the speakers demonstrate the same F_0 height *relationships* between competitive and noncompetitive incomings. This similarity across speakers suggests that they orient to a shared prosodic system for the management of overlap and turn competition. These findings represent an advance on the results presented by Shriberg *et al.* (2001a), who generally found raised F_0 in overlap onset, compared to non-overlap onset, but did not distinguish between competitive and noncompetitive overlaps. Our results indicate that raising F_0 is not a function of overlap *per se*, but of *competitive* overlap. However, this conclusion should be regarded as provisional, given the relatively small number of noncompetitive overlaps in the present data set.

- Do overlappees modify F_0 during competitive overlaps as opposed to noncompetitive overlaps?
- Are the F_0 modifications made by overlappees dependent on their interactional response to a competitive incoming?

It was found that when the speakers were overlapped noncompetitively they reduced their F_0 to below the typical F_0 height of their pre-overlap speech. Overlapped speakers did the same when not returning competition to a competitive incoming. These findings applied to each individual speaker as well as to the group. By contrast, the behaviour of speakers varied when responding competitively to a competitive incoming: two speakers routinely raised F_0, while the other three reduced it.

These findings on F_0 height in overlapped turns differ from previous findings reported by French and Local (1983). French and Local report no change in overlappees' pitch upon a noncompetitive incoming. Where overlappees yield the turn as a response to competitive incomings, French and Local (1983) make no mention of pitch but reported a decrease in volume. However, we found that overlappees marked both types of noncompetitive response by lowering F_0 compared to the immediately preceding stretch of speech. This seems to be a consistent strategy of all speakers.

Regarding return of competition by overlappees, French and Local (1983) found that it is signalled by decreased tempo and increased loudness from the point of the incomer's onset, which suggests that pitch does not play a major role for achieving this interactional goal. Our finding that return of competition does not always involve lowering of F_0, and that two of our speakers routinely raised F_0, suggests that F_0 is potentially a resource for marking the difference between competitive and noncompetitive responses to incomings. However, this seems to be a particular speaker's strategy rather than a general tendency. The ways in which speakers' individual F_0 modification interacts with other prosodic resources for turn competition is a topic for future investigation.

5. Conclusions

These results provide objective, quantitative support for earlier claims made on impressionistic evidence that participants in talk-in-interaction systematically manipulate F_0 height as part of the management of overlapping talk. Specific claims had been made by French and Local (1983) based on an analysis of a British English corpus. The present findings suggest that similar manipulations of pitch are used by American English speakers. It has long been acknowledged in intonation research that pitch height may have a phonological function, in the sense that choice of relatively high versus low pitch may convey differences in meaning (e.g. 'key' as described by Brazil, 1980). However, there have been few persuasive demonstrations of the meaning distinctions that speakers actually realise through pitch height variation in their spontaneous conversational interaction. A notable exception is Couper-Kuhlen's account of 'high onsets' in radio call-in interactions (Couper-Kuhlen, 2001). There is an opposition between low and high pitch on the initial stressed syllable of callers' turns that are potentially the 'anchor position' for the call, that is the sequential position in which the caller's reason for calling may be given. If a high onset is used, the caller's turn construction unit (TCU) will be treated as the reason for the call, and the radio host allows him/her to continue at length. By contrast, a low onset is treated as indicating that the TCU is not the reason for the call, but a preface to it, and the talk is routinely and swiftly followed by an intervention from the host. There may be some commonality between that use of high pitch and the use of high pitch in competitive incomings, described in the present study: in both situations, high pitch is associated with, and oriented to as, a claim for the floor; while low pitch is associated with noncompetitive talk that is not seeking an extended turn.

From a methodological perspective, the need to support impressionistic claims by instrumental measures has been recognised and applied in a variety of recent studies on prosody in interaction that use the method of CA (e.g. studies reported in Couper-Kuhlen and Ford, 2005). However, to the best of our knowledge this is the first study that applies this methodology to simultaneous talk. As outlined before, overlapping speech presents a challenge for this type of investigation as reliability of acoustic measurements is compromised in the inherently noisy situation of speaker overlap. Availability of single channel recordings on close talking microphones reduces the problem to a large extent, but still does not completely eliminate the influence of crosstalk. In addition, this recording set-up may limit the spontaneity of the discourse. The trade-off between naturalness of discourse and reliability of acoustic measures from noisy speech signal needs to be addressed, ideally by developing sophisticated signal processing techniques that are tailored for noisy rather than clean speech.

The findings of this study open up the possibility, admittedly still some way distant, of the automatic classification of instances of overlap as competitive or not, based on acoustic analysis. Such overlap models are potentially of use for automatic meeting transcription and can also inform development of more natural turn-taking strategies in human-computer dialogue systems. However, this paper has focused solely on the role of F_0 height in overlap management. It has been proposed that other phonetic features, such as intensity and speech rate may also be resources for overlap management, either as separate parameters or in combination. Investigation of this issue, which is the subject of our current research, is essential for further theoretical and technical progress in the study of simultaneous talk.

References

Baken, R. J. and R. F. Orlikoff (2000). *Clinical measurement of speech and voice*. San Diego, CA: Singular Publishing.

Brazil, D. (1980). *Discourse intonation and language teaching*. New York, NY: Longman.

Carletta, J. (1996). "Assessing agreement on classification tasks: The Kappa statistics". *Computational Linguistics* **22 (2)**: 249–254.

Caspers, J. (2003). "Local speech melody as a limiting factor in the turn-taking system in Dutch". *Journal of Phonetics* **31 (2)**: 251–276.

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction*. Amsterdam: John Benjamins.

Couper-Kuhlen, E. (2001). "Interactional prosody: High onsets in reason-for-the-call turns". *Language in Society* **30 (1)**: 29–53.

Couper-Kuhlen, E. and C. E. Ford (2005). *Sound patterns in interaction: Typological studies in language*. Amsterdam: John Benjamins.

De Cheveigné, A. (2006). "Multiple F0 estimation", in D. Wang and G. J. Brown (eds.), *Computational auditory scene analysis: Principles, algorithms, and applications*. Hoboken, NJ: IEEE Press/Wiley, 45–79.

De Cheveigne, A. and H. Kawahara (2002). "YIN, a fundamental frequency estimator for speech and music". *Journal of the Acoustical Society of America* **111 (4)**: 1917–1930.

Di Eugenio, B., P. W. Jordan, R. H. Thomason and J. D. Moore (2000). "The agreement process: an empirical investigation of human-human computer-mediated collaborative dialogs". *International Journal of Human-Computer Studies* **53 (6)**: 1017–1076.

Edwards, J. (2004). *The ICSI meetings corpus: Transcription methods*. Berkley, CA: International Computer Science Institute (ICSI).

French, P. and J. Local (1983). "Turn-competitive incomings". *Journal of Pragmatics* **7 (1)**: 17–38.

Galley, M., K. McKeown, J. Hirschberg and E. Shriberg (2004). "Identifying agreement and disagreement in conversational speech: Use of Bayesian networks to model pragmatic dependencies", *Proceedings of 42nd Meeting of the ACL*, Barcelona, Spain.

Janin, A., D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke and C. Wooters (2003). "The ICSI meeting corpus", *Proceedings of ICASSP* 2003. AU:1

Jefferson, G. (1973). "A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences". *Semiotica* **9 (1)**: 47–96.

Jefferson, G. (1983). "Two explorations of the organization of overlapping talk in conversation: Notes on some orderliness of overlap onset". *Tilburg Papers in Language and Literature* **28**: 1–28.

Jefferson, G. (1986). "Notes on 'latency' in overlap onset". *Human Studies* **9 (2–3)**: 153–183.

Jefferson, G. (2003). "A sketch of some orderly aspects of overlap in natural conversation", in G. H. Lerner (ed.), *Conversation analysis: Studies from the first generation*. Amsterdam: John Benjamins, 43–59.

Koiso, H., Y. Koriuchi, 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 (3–4)**: 295–391.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing. 5th Edition*. London: Academic Press.

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

Schegloff, E. (2000). "Overlapping talk and the organization of turn-taking for conversation". *Language in Society* **29 (1)**: 1–63.

Schegloff, E. A. (1982). "Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences", in D. Tannen (ed.), *Georgetown roundtable on languages and linguistics 1981; analyzing discourse: Text and talk*. Washington, DC: Georgetown University Press, 71–93.

Schegloff, E. A. (2001). "Accounts of conduct in interaction: Interruption, overlap and turn-taking", in J. H. Turner (ed.), *Handbook of sociological theory*. New York: Springer, 287–321.

Shriberg, E., R. Dhillon, S. Bhagat, J. Ang and H. Carvey (2004). "The ICSI meeting recorder dialog act (MRDA) corpus", *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, 97–100.

Shriberg, E., A. Stolcke and D. Baron (2001a). "Can prosody aid the automatic processing of multi-party meetings? Evidence from predicting punctuation, disfluencies and overlapping speech", ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding, Red Bank, NJ.

Shriberg, E., A. Stolcke and D. Baron (2001b). "Observations on overlap: Findings and implications for automatic processing of multi-party conversation", *Proceedings of EUROSPEECH 2001*, Aalborg, Denmark, 1359–1362.

Tamburini, F. and C. Caini (2005). "An automatic system for detecting prosodic prominence in American English continuous speech". *International Journal of Speech Technology* **8 (1)**: 33–44.

Wang, D. and S. Narayanan (2006). "An acoustic measure for word prominence in spontaneous speech". *IEEE Transactions on Speech and Audio Processing* **15 (2)**: 690–701.

Wells, B. and J. Corrin (2004). "Prosodic resources, turn-taking and overlap in children's talk-in-interaction", in E. Couper-Kuhlen and C. E. Ford (eds.), *Sound patterns in interaction*. Amsterdam: John Benjamins, 119–143.

Wells, B. and S. Macfarlane (1998). "Prosody as an interactional resource: Turn-projection and overlap". *Language and Speech* **41 (3–4)**: 265–294.

Wrigley, S. N., G. J. Brown, S. Renals and V. Wan (2005). "Speech and crosstalk detection in multi-channel audio". *IEEE Transactions on Speech and Audio Processing* **13 (1)**: 84–91.

Yngve, V. (1970). "On getting a word in edgewise". Papers from the 6th regional meeting, Chicago Linguistic Society, Chicago, IL, 567–578.

To access supplementary sound content for this chapter, go to http://intouch.emeraldinsight.com/. See page ? for details.

AUTHOR QUERY FORM

	Book: SP-V008 Chapter: 8	Please e-mail or fax your responses and any corrections to: E-mail: Fax:

Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

Disk use

Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article

☐ Scanning the artwork

Bibliography

If discrepancies were noted between the literature list and the text references, the following may apply:

☐ **The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.**

☐ ***UNCITED REFERENCES:*** **This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.**

Queries and/or remarks

Location in Article	Query / remark	Response
AU:1	Please provide the venue of the proceeding for refs. Janin *et al.* (2003), Shriberg *et al.* (2004).	