Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 1–23

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Resources for turn competition in overlapping talk

Emina Kurtić [a,b], Guy J. Brown [b], Bill Wells [a,*]

[a] Department of Human Communication Sciences, University of Sheffield, 31 Claremont Crescent, Sheffield S10 2TA, UK
[b] Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP, UK

Received 15 June 2012; received in revised form 21 September 2012; accepted 9 October 2012

Abstract

Overlapping talk occurs frequently in multi-party conversations, and is a domain in which speakers may pursue various communicative goals. The current study focuses on turn competition. Specifically, we seek to identify the phonetic differences that discriminate turn-competitive from non-competitive overlaps. Conversation analysis techniques were used to identify competitive and non-competitive overlaps in a corpus of multi-party recordings. We then generated a set of potentially predictive features relating to prosody (F0, intensity, speech rate, pausing) and overlap placement (overlap duration, point of overlap onset, recycling etc.). Decision tree classifiers were trained on the features and tested on a classification task, in order to determine which features and feature combinations best differentiate competitive overlaps from non-competitive overlaps. It was found that overlap placement features played a greater role than prosodic features in indicating turn competition. Among the prosodic features tested, F0 and intensity were the most effective predictors of turn competition. Also, our decision tree models suggest that turn competitive and non-competitive overlaps can be initiated by a new speaker at many different points in the current speaker's turn. These findings have implications for the design of dialogue systems, and suggest novel hypotheses about how speakers deploy phonetic resources in everyday talk.

Keywords: Overlapping talk; Turn competition; Prosody; Turn-taking; Turn end projection

1. Introduction

People do not usually talk at the same time. Conversations seem to be based on well-organised turn exchange systems, in which speakers take turns and cooperate to achieve overlap-free interaction, estimated to occupy around 90% of total speaking time (e.g. Shriberg et al., 2001b; Cetin and Shriberg, 2006). Simultaneous speech by two or more speakers is, nevertheless, frequently observed. If, rather than total speaking time, we consider the number of speaker turns that are overlapped, the incidence of overlapping talk is much higher. For example, Heldner and Edlund (2010) estimate that 41–45% of all turn shifts between speakers in spontaneous conversational dyads contain overlap, and Shriberg et al. (2001b) report that 30–50% of all turn exchanges in multi-party meetings contain some overlap. This raises a number of questions about the status of overlapping speech in turn-taking: Why does overlap occur with such frequency? Is it an integral part of the turn-taking system, a by-product of otherwise one-speaker-at-a-time turn exchange? Or is it a conversational tool used by speakers to achieve certain communicative goals?

Most previous studies on turn taking and speaker overlap at least allow for the latter possibility, agreeing that overlapping talk is an environment in which *turn competition* may take place. It follows that some instances of speaker overlap will be turn competitive, while other overlaps will be non-competitive. This observation raises the question that the current study seeks to address: *If overlapping talk is the domain of different communicative actions such as competing vs. not competing for the turn, how do conversation participants display these differences to one another?* An answer to this question would enhance our understanding of how people deploy phonetic and linguis-

* Corresponding author. Tel.: +44 (0) 114 222 2429; fax: +44 (0) 114 273 0547.
E-mail address: bill.wells@sheffield.ac.uk (B. Wells).

0167-6393/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2012.10.002

tic resources in everyday talk, enabling us to address a number of important theoretical and practical questions. Are there interactional 'universals' in the management of overlapping talk or is it language (or culture) specific (c.f. Sidnell, 2001)? How might an answer to this question contribute to the study of intercultural communication? How do young children learn to manage turn-taking in general, and overlap in particular (c.f. Wells and Corrin, 2004)? What light might this shed on the interactional problems of individuals with communication difficulties, arising for example from autism or hearing loss?

An answer to our question might also contribute to improvements in speech technology. Reidsma et al. (2011) claim that differentiating between turn competitive and non-competitive overlapped incomings is an essential part of so-called 'continuous conversation' with a virtual agent. An automatic dialogue system needs know when to yield the turn to the human user, which also involves being able to deal with the cases when the human user takes the turn while the system is still talking. To achieve this, the system has to be able to recognise such incomings as turn-competitive and employ practices for management of turn-competitive incomings. On the other hand, the dialogue system should also be able to produce non-competitive overlaps such as response tokens (backchannels) at appropriate places to acknowledge receipt of the ongoing turn (Gravano and Hirschberg, 2011). Findings on the organisation of human overlap management, and in particular on differentiating between turn-competitive and non-competitive overlaps, could thus be a particularly important source of knowledge for automatic systems that aim at spontaneous conversation with human users.

The focus of the present study is solely on the acoustic and temporal features of overlap. We make no claims about participants' use of non-verbal cues in the realisation of turn competitiveness since, for reasons given in Section 3 below we chose to work with the ICSI meeting corpus, for which only audio recordings exist. The role of gesture for conversational sequencing and the structuring of turn-taking has long been recognised and analysed (e.g. in Goodwin, 1980; Goodwin and Goodwin, 1986; Kendon, 1967; Bavelas et al., 2002; Barkhuysen et al., 2008). However, there has been little research specifically concerned with the relevance of non-verbal cues for turn competition in overlap, with the exception of two recent studies which support the view that gestures are relevant resources for overlap management in face-to-face discourse. Lee et al. (2008) show that adding hand movements to intensity analysis improves discrimination between turn-competitive and non-competitive overlaps in their corpus of acted scripted dialogues. In a study of French mundane conversations Mondada and Oloff (2011) show that continuing vs. abandoning gesturing during overlap is associated with how problematic participants take the overlap to be. These studies indicate that the role of gesture and gaze in relation to phonetic features in overlapping talk is a promising area for future research. However, it will be dependent on access to corpora where individual speakers are recorded on separate channels and where the video recordings provide sufficient detail of each participant's gesture and gaze behaviour (e.g. Carletta, 2007; Kurtic et al., 2012).

The methodology of the present study draws on complementary traditions of research into overlapping talk: speech science, conversation analysis and interactional phonetics. First, we review the contribution of each of these traditions to the study of overlap. On the basis of that research, we identify a set of temporal, prosodic and other features that may be implicated in the design of overlapping talk. We describe how we constructed a collection of overlaps from naturally occurring, unscripted multi-party meetings and how these were classified as competitive or non-competitive. Decision tree analyses are then used to identify the role of prosodic and non-prosodic features in differentiating competitive from non-competitive overlaps. The resulting decision tree models enable us to make a number of empirically grounded, testable hypotheses about how human participants signal competition for the turn. Finally, we explore the theoretical and practical implications of our hypotheses.

2. Traditions of research into overlapping talk

2.1. Speech science research into overlapping talk

As indicated above, the speech technology community has an interest in understanding more about overlapping talk, in order to improve spoken dialogue systems for example. This has fuelled research into the acoustic and temporal properties of overlap. Shriberg et al. (2001b) carried out a quantitative study of overlaps from the ICSI corpus, described below. The study is fairly typical of speech science research into overlapping talk, in that the analysis is conducted on a large corpus of audio recordings of more or less naturalistic spoken interaction. Each speaker was recorded onto a separate audio channel by a close-talking microphone, thus allowing for acoustic analysis of individual speakers talking in overlap. The raw acoustic data was analysed to search for recurrent acoustic correlates of overlap. Fundamental frequency (F0) and energy at the onsets of turns in overlap were found to be high, compared to the onsets of turns from silence (i.e. not in overlap). However, the study did not differentiate between competitive and non-competitive overlaps. For this reason, the conversational function of these prosodic resources remained unclear.

More recently, Gravano and Hirschberg (2011) have analysed the Columbia Games Corpus in order to identify the prosodic, syntactic and acoustic cues that precede turn changes, turn retentions and backchannels. They found that inter-pausal units (IPUs) that precede turn transitions with and without overlap exhibit comparable turn-yielding cues. However, their study only considered smooth turn changes, and therefore did not address cues that potentially signal competitive and non-competitive overlaps. Indeed,

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002

turn-competitive overlaps were specifically ignored because "they correspond to disruptions of the conversational flow at arbitrary points during the speaker's turn, rather than unobtrusive overlap during fluent exchanges." (p. 626).

With regard to temporal features of overlap, Heldner and Edlund (2010) measured overlap duration in three very large corpora. In line with earlier research, they found that many turns are overlapped: over 40% show an overlap of longer than 10 ms. They report in some detail on the duration of the between speaker intervals, observing that 70–82% of these, including both gaps and overlaps, were shorter than 500 ms, and that the maximum length of overlap seems to be around 3 s. It is thus clear that there are strong statistical tendencies in speakers' behaviour with regard to overlapping talk. However, Heldner and Edlund (2010) offer little in terms of a systematic explanation for these tendencies in terms of speaker behaviour. Indeed, they appeal to "anecdotic (sic) data and introspection" rather than to research findings when discussing possible explanations for the distribution of overlap:

> "*For example, overlapping the end of a highly predictable utterance may be entirely acceptable, whereas overlap into completely unpredictable content may be disturbing or rude. Whether the predictability of an utterance and its speech act are key factors remains to be investigated* ..." (p. 565).

However intuitively appealing such speculations may be, it is important for the scientific study of talk that we seek empirical support for an explanation of how overlap works. For this, we turn to a different tradition of research into overlapping talk, which has explored how overlapping incomings are constructed by the new, incoming speaker and how they are responded (or oriented) to by the current, overlapped speaker.

2.2. Conversation analysis research into overlapping talk

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 the projectability of turn-ends. Self-selection occurs in cases where the current speaker does not select the next speaker, so that when the current speaker reaches a TRP, 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 in progress, but at the point where such completion is projected, giving rise to so-called *terminal overlaps*. 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.

In addition to non-competitive overlaps at the TRP, i.e. the terminal overlaps and simultaneous starts described above that arise as a by-product of the turn-taking system, a common type of overlap is the so-called continuer (Schegloff, 1982), backchannel (Yngve, 1970) or *response token* (Gardner, 2001; Stivers, 2008), routinely used by overlappers to mark receipt of the ongoing turn and confirm the current speaker's right to an extended turn. Two further types of non-competitive overlaps have been described by Lerner (1999a,b): *collaborative completions*, where the incoming speaker overlaps and completes the turn started by the current speaker; and *choral productions*, where two or more participants produce for example a greeting or a toast in overlap.

Several subsequent studies, e.g. Jefferson (1983) and Schegloff (2000), contrast such non-competitive overlaps with those in which participants compete for the turn in progress. French and Local (1983) define turn competitive incomings in overlap 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) characterizes such 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: according to Schegloff (2001), attempting to 'drive the prior speaker out' can be the aim of either party.

In the course of identifying competitive and noncompetitive overlaps and their various subtypes, conversation analysts have described some of the linguistic resources employed by conversational participants in order to display an overlap as turn competitive or as noncompetitive. 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 Jefferson, the positioning of the overlap onset is related to the competitiveness of the overlap. She offers a preliminary categorisation of overlap onsets into *transitional*, *progressional* and *recognitional* onsets, according to their position relative to the TRP. 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, by contrast, are located at points where the incoming speaker has gained sufficient understanding of the current speaker's turn. Heldner and Edlund (2010) also mention such cases:

> "*Many overlaps occur because the next speaker is confident about what the current speaker will say, and deliberately responds before the current speaker finishes. Speaker changes often occur when the current utterance becomes predictable in the eyes of the next speaker, so*

with respect to timing, projection of content may result in overlaps . . ." (p. 566)

These onsets result in "first-order overlaps of varying degrees of turn incursion" (Jefferson, 1983), p. 28. Jefferson's "byproduct" and "first-order" overlaps thus correspond to non-competitive and turn competitive overlaps respectively.

In addition to the placement of the overlap onset, Jefferson identified certain *temporal* features, relating to the extent and positioning of the overlap, as possible resources for turn competition. For example, Jefferson (2003) reports that if a speaker does not drop out of the overlap, but continues whilst being aware that overlap is taking place, this is associated with turn competition. Schegloff (2000) notes a series of features that are deployed by speakers in the course of competitive overlap. These include speech rate, cut-offs, sound stretches and repetition or recycling of prior material.

2.3. Interactional phonetic research into overlapping talk

A related line of research has adopted a Conversation Analysis approach to the study of interaction, with its emphasis on corpora of naturally occurring talk, sequential analysis and participant orientation as a primary source of evidence, while adding a more sophisticated approach to phonetic analysis. This type of work has been called 'Interactional Phonetics', the principles of which are set out by Local and Walker (2005). Several studies (French and Local, 1983; Couper-Kuhlen, 1993; Wells and McFarlane, 1998; Schegloff, 2000; Lee et al., 2008; Kurtić et al., 2009) have claimed that *prosodic features*, including fundamental frequency height, intensity, speech rate and rhythm are important resources for turn competition in overlap.

In one of the first interactional phonetic studies, French and Local (1983) proposed that it is the combination of raised pitch and volume that fulfills this function. They offer evidence that this combination is utilised by overlapping speakers (henceforth, *overlappers*) to compete for the turn, and is also treated as competitive by turn-holders (henceforth, *overlappees*). French and Local (1983) suggest that the timing of 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 its pragmatic function (i.e., being an agreement or disagreement) as being robust features for discrimination between competitive and non-competitive overlaps. Pitch and volume have subsequently been reported in connection with overlap management by Shriberg et al. (2001a), and by Schegloff (2000), who regards increases in pitch or volume as turn competitive "hitches" that occur in competitive overlaps.

The relationship between positioning of overlap onset and prosodic design of the incoming was investigated by Wells and McFarlane (1998). Synthesising the analyses of French and Local (1983) and Jefferson (1983), they claim that the combination of raised pitch and loudness 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 McFarlane, 1998, p. 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 raised pitch and loudness, in which case they are not treated as turn-competitive despite their placement.

While interactional phonetic research has resulted in suggestive accounts of how overlapping talk is designed and how it is used by conversational participants, these studies are in various ways restricted in their scope. Some focus on particular overlap types. For instance, French and Local (1983), Schegloff (2000) and Kurtić et al. (2009) only consider overlaps placed clearly prior to possible completion. In some studies, only a subset of prosodic features is analysed. For instance, French and Local (1983) and Wells and McFarlane (1998) only considered pitch and loudness; Lee et al. (2008) intensity; Kurtić et al. (2009) fundamental frequency; Couper-Kuhlen (1993) speech rhythm and Kurtić et al. (2010) speech rate. Consequently, these studies cannot explain how all these features might interact. They thus offer only a partial insight into how turn competition in overlap works.

With the aim of providing a more comprehensive account of the phonetics and phonology of overlap than is available to date, we investigate the distribution of prosodic and positional features, as well as their combinations, that are used by speakers in competitive and non-competitive overlaps. We attempt to remain as open as possible with regard to hypotheses about how these features work. For example with regard to the possible locations of turn competitive incomings, Wells and McFarlane (1998) and Schegloff (2000), following Jefferson (1983), exclude the possibility of turn-competitive overlaps around the possible endings of the ongoing turn. A priori they limit their consideration of turn competition to cases in which overlappers come in clearly prior to the overlappee's turn completion. However, in the current study we allow for the possibility that an overlapping incoming in terminal position may sometimes be competitive.

If this is so, then based on this previous work we can expect that turn competitive overlaps will be positioned turn-incursively, i.e. well before a possible projected TRP, while non-competitive overlaps will mainly be positioned around possible turn completions. In addition, we hypothesise that this positioning has a bearing on the prosodic design of the overlaps. To compete for the turn in an environment where this would not be expected, i.e. at a TRP, may require different resources than it would when the incoming clearly violates the current speaker's right to the turn.

Many of the limitations of previous approaches derive from methodological difficulties associated with interactional phonetic work on overlapping talk. Carrying out phonetic analysis, whether auditory/perceptual or acoustic/instrumental, of overlapping talk in recordings of naturally occurring conversations is very difficult because of the

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002

Table 1
Amount of overlap in our corpus, drawn from the ICSI data set (including both two-speaker overlaps and multi-speaker overlaps). Dialog act segments were derived by Dhillon et al. (2004) based on syntactic, pragmatic and prosodic criteria.

Total speaking time (hh:mm:ss)	Total overlap time (% of total speaking time)	Total number of segments	Overlap instances (% of total number of segments)
05:24:32	11.53	9816	45.6

problem of differentiating the signals from the overlapping speakers. While the temporal onset and offset of overlap are relatively straightforward to identify, the extraction of features such as F0 and intensity poses a major challenge for current sound separation techniques, as it does for the listener, however skilled. The method employed in the present study aims to combine the strengths of the 'speech science' and 'interactional phonetics' approaches, in order circumvent the methodological limitations of each.

3. Materials and methods

We employ a methodology in which acoustic and temporal features, including fundamental frequency, speech intensity, speech rate and pausing, are extracted from a large collection of turn-competitive and non-competitive overlaps. A machine learning technique – decision tree modelling – is then applied to analyse the relationship between these features and turn competition. This methodological approach differs from previous interactional phonetic studies in terms of the size of the collection of overlaps and by using machine learning methods to investigate the relationship between the features and turn competition. However, the methods of interactional phonetics are used to build the collection of overlaps and to derive the features to be studied. This departs from the annotation practices usually used in dialogue modelling (e.g., Dhillon et al., 2004). On this basis we are able to make empirically grounded hypotheses about the relevance of prosodic and non-prosodic parameters as turn-competitive resources used by the participants themselves, while at the same time exploiting the statistical advantages of a large data set.

3.1. Building the collection of overlaps

We base our analyses on a subset of the ICSI Meeting Corpus (Janin et al., 2003) comprising eight meetings.[1] In selecting our data subset, we aimed to control for speaker, meeting type and, as far as possible, for the number of meeting participants. The meetings we selected include five (three male and two female) speakers of American English (AE) as a first language, who are the most frequent participants in all the meetings. By tracking these speakers across several meetings we could obtain enough acoustic data for our analytical purposes. The meeting set also includes two native AE speakers whose speech is included in the analysis, but who are far less talkative than the five selected speakers. The decision to restrict analysis to the five speakers represents a compromise between the desire to constrain the amount of speaker-specific variability arising from accent, age etc., while including sufficient speakers to capture practices of turn taking and overlap that are shared across this speech community. Each participant in the ICSI meetings is recorded on a separate audio channel at a sample rate of 16 kHz with 16 bit resolution. To minimise the possibility of variations in microphone placement, we chose speakers who used headset rather than lapel microphones. Nevertheless, there is likely to be some variability of sound level due to microphone placement.

Another criterion in selection of the data subset was to use meetings of the same type. As reported in previous studies based on the same corpus (Shriberg et al., 2001b), some ICSI meetings are more spontaneously interactive, while others are directed by one person, with other participants reporting on their work in turn. The meetings that we selected are all of the spontaneous type with six, seven or eight participants including the five selected speakers. Other studies, such as Heldner and Edlund (2010) also use data gathered from the popular 'map task' scenario, where the main function of the interaction is to provide recorded data for research purposes. However, researchers in Conversation Analysis and Interactional Phonetics try to limit the data they use to recordings of participants conducting their ordinary business, be that a conversation between friends, a medical consultation or a research meeting, as in the present case.

Overlap instances were detected automatically using the start and end time information for each segmentation unit and for each word within that unit. The basic segmentation unit we use is the *dialogue act segment* or simply *segment*. Segments are units which are determined by a hand labelling procedure for dialogue acts in the ICSI Meeting Corpus, described in detail by Dhillon et al. (2004).[2] This segmentation is based on syntactic, pragmatic and prosodic criteria, so segments resemble turn constructional units (TCUs), which have been constituted as basic turn constructional resources (Sacks et al., 1974) and subsequently widely considered in conversation analysis research (Selting, 1998).

Each segment is associated with a start and end time that aligns it with the speech recording. 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 for meeting data (Hain et al., 2012). For overlap detection, first, all segments that contained overlaps were identified by their start and end times. Then, word-level forced alignments of the corpus were used to identify which words

[1] The meeting designations are Bmr006, Bmr007, Bmr008, Bmr013, Bmr016, Bmr018, Bmr022 and Bmr025.

[2] This data was downloaded from <http://www.icsi.berkeley.edu/~ees/dadb/>.

overlapped within a segment. The whole overlap region was then delimited by the start time of the first overlapping word and end time of the last overlapping word.

We expected the quality of our forced alignments to be similar to those obtained by Kurtic et al. (2012), who used the same speech recogniser to force-align spontaneous conversation between friends in British English. They reported a 20 ms error (i.e. the proportion of boundaries placed more than 20 ms away from the ground truth boundary) of 35%. However, manual random checks found that the forced alignment error was generally much lower in the current study, for two reasons. Firstly, Kurtic et al. (2012) report that misalignment was mainly found in cases where laughter or long outbreaths were overlaid on speech or in regions of whispered or creaky voice. These phenomena are much more frequent in the spontaneous conversation between friends that they used, rather than in the meeting talk used here. Secondly, the speech recogniser used in both studies was trained on multi-party meetings; it therefore performs better on forced alignment of meeting data than on the spontaneous conversational data of Kurtic et al. (2012).

The total amount of overlap in the data set is shown in Table 1. Two types of overlaps have been identified in multi-party data, as shown in Fig. 1: two-speaker and multi-speaker overlaps. *Two-speaker overlaps* describe two different scenarios. The first is the case in which a single overlapper overlaps the current turn holder (overlappee). The second is the case in which two overlappers overlap the same turn held by an overlappee; however, their incomings are placed at different times in the overlappee's turn and do not overlap each other. *Multi-speaker overlaps* are overlaps in which at least one word from the overlappee's turn is overlapped by multiple overlappers. Multi-speaker overlaps are common, but generally very short: in our data, around 30% of all overlap instances involve multiple speakers, but only 2.8% of the entire in-overlap time results from multi-speaker overlap. Because it is possible that multi-speaker overlaps have distinct, as yet unknown characteristics in terms of their timing and linguistic/phonetic design, we decided to exclude them from analysis in the present study. Only two-speaker overlaps are therefore considered in the following.

3.2. Competitiveness annotation

Competitive overlaps are those in which either or both speakers demonstrate that they want the turn for themselves at that very moment, and not when the other party has completed his/her turn. These overlaps are treated as problematic by the overlapping speakers, and potentially also by other speakers participating in conversation prior to or after the overlap in question. Whether an overlap is competitive or not was established by analysing the conversational sequence in which overlap occured.

For example, consider the overlap in line 6 of the conversation extract shown in Extract (1).[3]

(1) ICSI_Bmr018_566:

```
1   m13: we (0.8) let's try that again
2   f16: [yes ]
3   f08: [yeah] (0.2) that's good
4   m11: [OK  ]
5   f16: [so  ] and maybe we won't
         [laugh this time also]
6 > m11: [so remember        ]
         to read (.)the transcript
         number (0.2) so that uh everyone
          knows that what it is and (1.0)
         ready three two one
```

Male speaker *m11* starts his turn in line 6 at a point in female speaker *f16*'s turn that is not a point of syntactic completion. *m11* continues past the point where *f16* stops and neither *f16* nor any other speaker attempts to regain the floor. In his competitively incoming turn, *m11* has introduced a new subtopic: the participants are preparing to read out aloud a transcript of an earlier meeting, which has been a problematic activity for them in the past. *m11*'s new subtopic is to remind them of a point of recording procedure.

A further example of a competitive incoming is seen in line 7 of Extract (2):

(2) ICSI_Bmr006_158:

```
1   m11: um, I had a I spoke with some
         people up at Haas Business
         School who volunteered
2   m11: should I pursue that
3   f16: oh definitely [yeah]
4   m11:               [yeah]
5   m13: [yeah]
6   m11: [so  ] they they originally
         (0.2)
         they've decided not to do go
         into speech (0.3)
         so I'm not sure whether they'll
         still be so willing to
         volunteer but I'll
         [send an email and ask]
7 > m13: [tell them about the free lunch]
8   m11: I'll tell them about the free lunch
9   f16: yeah (0.2) [yeah]
10  m11:            [and] they'll say
         there's no such thing (0.5) so
11  f16: I'd love to get people that are
         not linguists or (.)
         [engineers]
12  m11: [right    ]
         (0.6)
13  f16: cuz (0.2) these are both (.) weird
14  m11: right
```

[3] See Appendix for transcription conventions.

m13 starts, in line 7, at a point well before *m11*'s turn is syntactically complete. As in Extract (1) above, both speakers, talking in overlap, reach a point of completion of their turn. Unlike in (1), in (2) it is the overlappee (*m11*), rather than the overlapper (*m13*), who takes the subsequent turn (l. 8). However, that he aligns with *m13*'s new subtopic of the 'free lunch' shows that *m13* has nevertheless succeeded in redirecting the topical content of the talk by competing at line 7.

A third and final example of turn-competition is seen in Extract (3).

(3) ICSI_BMR007_109:

```
1 m13:   well but see I find it
         [interesting]
2 f16:   [so:    ]
3 m13:   even if it wasn't any more
         (0.2) because (.) since we were
         dealing with this full
         duplex sort of thing in
         Switchboard where it was
         just all separated out .hhh
4 f16:   mm-hmm
5 m13:   we just everything was just
         nice so the (.) so the issue
         is in (.) in a situation
         (0.4) [where tha that's    ]
6 > f16: .hhh [well it's not really]
         (.) nice it depends
         what you're doing
         so if you were actually
         .hhh (0.4) having (0.3) uh
         (0.5) depends what you're
         doing if (1.2) right now we're
         do we have individual
         mics on the people in this
         meeting
7 m13:   mm-hmm
```

Female speaker *f16* starts her turn in line 6 at a point in male speaker *m13*'s turn that is not a point of syntactic completion. Even though *f16* starts the overlap during the second part of *m13*'s turn (beginning with "so the ... so the issue is in ..."), she chooses to address the first part of it ("everything was just nice"), attempting to bring the topic back to "nice" and thereby preventing *m13* from continuing towards turn completion. *m13* abandons his turn: unlike the overlappees in (1) and (2), he breaks it off before reaching a point of syntactic completion. Thereupon, *f16* secures the floor for an extended turn despite her many disfluencies and long pauses, no other participants attempt to take over from *f16*. These positional, syntactic and pragmatic criteria offer evidence of *f16*'s turn-competitive behaviour, while *m13* withdraws from the overlap.

The non-competitive class consists of overlaps that are not treated as problematic by participants. Some overlaps, like response tokens, choral productions and collaborative completions, are considered to be generally non-competitive (Schegloff, 2000). However, there are also overlaps which do not belong to one of these categories, but in which participants nevertheless do not display evidence of turn competition. We call these overlaps *other non-competitive overlaps*. Several examples of other non-competitive overlaps are given in Extract (4).

(4) ICSI_Bmr007_271–275:

```
1   ml1: so if you fiddle around with
         it a little bit and you get
         good numbers you can actually
         do a pretty good job of
         segmenting when someone's
         talking and when they're not
         but if you try to use the same
         parameters on another speaker
         it doesn't work anymore
         even if you normalize it based
         on the absolute loudness
2   fl6: but does it work for that one
         speaker throughout the whole
         meeting
3   ml1: it does work for the one
         speaker throughout the whole
         meeting um (0.7) pretty well
  >      [pretty well]
4 > ml8: [how     ] did you do it
         Adam
5   ml1: how did I do it
  >      [what do you] mean
6 > ml8: [yeah] (0.4) I mean wh
         what was the
7   ml1: the algorithm [was] uh (0.7)
8 > ml8:               [yeah]
9    ml1: take (0.5) o(.) every frame
          that's over the threshold
          (0.3) and then median-filter
          it (0.7) [and]
10   ml8:          [mm-hmm]
11   ml1: then look for runs
          so there was a minimum run
          [length so that]
12 > ml8: [every        ] frame
          that's over what threshold
13   ml1: a threshold that you pick
14   ml8: in terms of energy
15   ml1: yeah
```

In line 4, *m18* apparently predicts the end of *m11*'s turn and poses a request for clarification that occurs simultaneously with *m11*'s added increment "Pretty well". *m11* does not treat this overlap as competitive as he orients to *m18*'s question by reiterating its contents in line 5 to

ask for clarification himself. *m18* does not treat this as a clarification request, though, but more likely as a question that reiterates his in-overlap speech which may have passed unheard. This is indicated by his "Yeah" response, which itself gives rise to a non-competitive overlap (ll. 5–6). This overlap is a simultaneous start at a completion point at which *m11* reveals that his question from line 5 is a clarification request, not a repetition of unheard in-overlap talk. This overlap is thus a consequence of two speakers trying to solve an interactional problem (misunderstanding) in which they co-operate rather than compete.

The next overlap in line 8 follows a collaborative completion of *m18*'s talk (l. 7) by *m11* which itself does not involve overlap. "The algorithm" completes *m18*'s preceding turn. According to Local (2005), after a collaborative completion the original speaker will resume the turn; this is exactly what happens in line 9. *m18* confirms that the completion is correct and displays no intention of continuing to talk, which he also signals by the response token in line 10 that confirms *m11*'s right to speakership. *m11* obviously plans to continue his turn (l. 7); however, upon overlap by *m18*, he utters a filled pause ("uh") followed by a silence of 0.7 s before continuing. In this way he displays that he is waiting for *m18*'s further action and acknowledges *m18*'s right to the turn at this point. There is no obvious sign of turn competition in this short overlap.

The final example of other non-competitive overlap in this extract is in line 12. This starts at a potentially syntactically-complete point which *m18* may have projected as the end of *m11*'s turn. *m18*'s incoming contains again a request for clarification in the form of a question, as indicated by *m11*'s response (l. 13). Although *m11* continues past the point of potential completion after "run" in line 11, he drops out of overlap upon *m18*'s first word, i.e. either at or very close to the realisation point that there is overlap. In his subsequent talk (l. 13) he does not thematically continue his turn from line 11, but instead answers *m18*'s question: their exchange continues smoothly until the clarification is achieved. In this way *m11* demonstrates that he is not treating *m18*'s incoming as competitive but rather acknowledges *m18*'s right to the turn at that point.

All overlap instances in the data set were categorised into competitiveness classes in this way. Table 2 shows the count of two-speaker overlaps in the data set and their distribution across the competitiveness categories. To assess the inter-annotator agreement on competitiveness classification, two additional annotators categorised a subset of 785 overlaps from one randomly selected meeting in the same manner. The agreement between the three annotators was measured using Krippendorff's α coefficient. The overall inter-annotator agreement on competitive/non-competitive classification was $\alpha = 0.62$, comparable to that reported by Adda-Decker et al. (2008) for a similar annotation task.

Following Schegloff (2000) we excluded response tokens, choral and collaborative productions from later analyses and only considered the set of "other" non-competitive overlaps. As Schegloff has established, response tokens, choral and collaborative productions are mostly non-competitive by their interactional properties alone, i.e. their usage within the conversational sequence is such that it does not lead to turn competition. However, there is an appreciable amount of overlaps which do not implement turn competition but where the reason for their non-competitiveness is not immediately evident. In other words, in order to be understood as non-competitive, these overlaps need to be designed as non-competitive, for example by using a set of features such as the ones studied here. Therefore, we use this set of "other" non-competitive overlaps for studying the features that discriminate between competitive and non-competitive overlaps. When considering only "other" non-competitive overlaps, the inter-annotator agreement on competitive/non-competitive classification reduces to $\alpha = 0.56$. The final set of overlaps used for analysis contains 1455 overlap instances, after excluding some in which the audio quality was insufficient: 703 (47.4%) turn-competitive and 752 (52.6%) non-competitive.

3.3. Prosodic features

We analyse the following feature groups as prosodic resources for turn competition: fundamental frequency, speech intensity, speech rate and pausing features.

3.3.1. Fundamental frequency (F0), speech intensity and speech rate (SR)

We compute features that describe the distribution and dynamics of F0 and intensity over units of time (specified in Section 3.3.1.2). For example, for F0 we compute the mean, range and standard deviation over a time window,

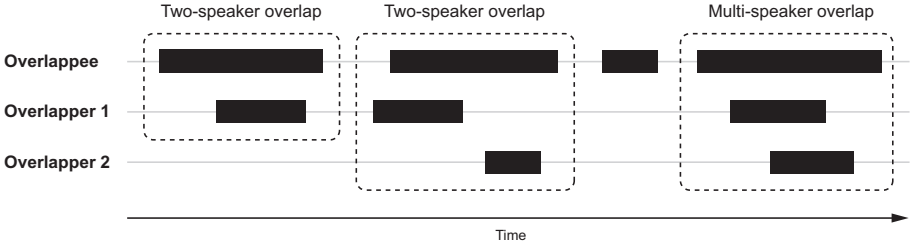

Fig. 1. Types of overlap in the multi-party recordings, showing two examples of two-speaker overlap and a multi-speaker overlap.

Table 2
The distribution of overlap instances across competitiveness categories for two-speaker overlap. 41 overlap instances (6 competitive and 35 other non-competitive) were excluded from the analysis because the speech signal was not of sufficient quality due to interference from other sounds, microphone errors, etc.

Competitiveness		Count	% of all overlap instances
Competitive		709	28.8
Non-competitive	Response tokens	807	32.8
	Choral productions	112	4.6
	Collaborative completions	46	1.8
	Other non-competitive	787	32.0
Total		2461	100.0

together with quantities that characterise the F0 contour. Praat (Boersma, 2001) was used to extract the F0 and intensity contours.

Speech rate (SR) is measured as the number of consonant-vowel (CV) intervals per second within the chosen time unit. Our decision to use CV intervals instead of syllables follows Dellwo et al. (2006), who express speech rate in terms of CV intervals per second, instead of syllables per second. They claim that this approach gives a more objective measure in fast speech (Dellwo et al., 2006). The CV intervals were computed from phone level transcriptions of the ICSI meetings obtained via forced alignment (Hain et al., 2012). We consider SR to be the articulation rate, i.e. pauses were excluded from the computation of SR, in order to measure whether there is a slow-down/speed-up in the speaker's speech.

The individual F0, intensity and speech rate features are shown in Table 3. For simplicity, Table 3 only shows the general names of the features. In practice, each of these features is computed for both overlapper and overlappee in each overlap, and is furthermore compared across different contexts and computed over different time units, as explained below. The full name of each feature as used in the results and discussion sections has the following form:

< `er` for overlapper | `ee` for overlappee >< *featurename* >< *context* >< *timeunit* >

For example, the feature denoted `erFOmeanInReClearWord5` encodes an overlapper's mean F0. The context is `InReClear`, indicating that the mean F0 is computed over the region of overlapping talk and measured relative to that speaker's mean F0 in clear turns (i.e., turns without overlap). Finally, the time unit `Word5` indicates that the feature is computed over the first 5 words of the overlap.

F0 and intensity contours are compared to contours derived from other contexts (such as clear segments or pre-overlap talk, as described below). We aim to measure the similarity in F0 and intensity slopes between successive time frames, and thus describe the similarity between the in-overlap contour and that found in the other context. For this we first compute the in-overlap gradient between pairs of successive F0 or intensity values as:

$$g_{f_n} = \frac{t_{n+1} - t_n}{f_{n+1} - f_n} \tag{1}$$

where n is the frame index and f represents the F0 or intensity value at that frame. We then compute the similarity between vectors that represent the in-overlap contour and those representing the contour of the other context using cosine similarity, which measures the similarity between two vectors by finding the cosine of the angle between them:

$$sim(\mathbf{v}_1, \mathbf{v}_2) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\|\|\mathbf{v}_2\|} \tag{2}$$

Here, $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors containing gradient values $\{g_{f_1}, g_{f_2}, \ldots, g_{f_M}\}$ where M is the number of time frames over which the similarity is computed, which in turn is determined by the unit of time that the feature spans (see below). Finally, the contour similarity feature is computed by averaging over the entire data set. For example, the fol-

Table 3
Fundamental frequency, intensity and speech rate features. All features were measured over a variety of contexts (Section 3.3.1.1) and time units (Section 3.3.1.2).

Feature name	Description
F0, INT and SR features	
`FOmean`	A speaker's mean F0 (Hz)
`FOSD`	The standard deviation of a speaker's F0 (Hz)
`FOrange`	The F0 range (maximum–minimum) of a speaker's talk (Hz)
`FOcontourSim`	The average similarity between a speaker's in-overlap F0 contour and their F0 contours produced in another context
`intMean`	A speaker's mean intensity (dB)
`intSD`	The standard deviation (SD) of a speaker's intensity (dB)
`intRange`	The intensity range of a speaker's talk (dB)
`intContourSim`	The average similarity between a speaker's in-overlap intensity contour and their intensity contours produced in another context
`SR`	A speaker's speech rate (consonant-vowel intervals/second)

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002

lowing computes the similarity between the overlapper's in-overlap F0 contour and the same speaker's F0 contours in clear turns:

$$\texttt{erF0contourSimInReClear} = \frac{1}{N}\sum_{k=1}^{N} sim(\mathbf{v}_{in}, \mathbf{v}^{k}_{clear}) \quad (3)$$

where $\mathbf{v}_{in}$ is the vector of F0 gradient values for the speaker's in-overlap speech, $\mathbf{v}^{k}_{clear}$ is the vector of F0 gradient values for the same speaker's kth clear segment and N is the number of clear segments for that speaker in the corpus.

3.3.1.1. Contexts. If F0, intensity and SR are used as turn-competitive resources, then we expect them to be used differently in competitive overlaps than elsewhere in the conversation. Therefore, features from regions of overlapping talk are compared against features obtained from different contexts, as described below and shown diagrammatically in Fig. 2.

In-overlap talk compared to clear talk (`InReClear`). For overlappers, a z-score is computed for the overlapper's in-overlap features (F0, intensity and SR) relative to the mean and standard deviation of the overlapper's features in clear turns. For example, for the F0 mean feature:

$$\texttt{erF0meanInReClear} = \frac{F0_{er,in} - \mu_{er,clear}}{\sigma_{er,clear}} \quad (4)$$

Here, `erF0meanInReClear` represents the normalised value of the mean fundamental frequency $F0_{er,in}$ computed for the overlapper's in-overlap talk. The terms $\mu_{er,clear}$ and $\sigma_{er,clear}$ denote the mean and standard deviation of the overlapper's F0 computed in clear turns. In this way we obtain an indication of how many standard deviations the in-overlap feature values are above or below the speaker's mean feature values in clear turns. Analogous features for intensity and SR are computed in the same way. For F0 and SR, clear turns from the entire data set of 8 meetings are taken for this normalisation. All intensity normalisations are done per meeting because the microphone type, position and amplification gain vary across meetings.

Similarly, the overlappee's in-overlap talk is compared to the overlappee's clear turns. By analogy with `erF0meanInReClear`, this is computed for the F0 mean feature as follows:

$$\texttt{eeF0meanInReClear} = \frac{F0_{ee,in} - \mu_{ee,clear}}{\sigma_{ee,clear}} \quad (5)$$

In-overlap talk compared to overlappee's pre-overlap talk (`InRePre`). For an overlapper, this is computed as a difference between z-scores (relative to the clear turns) of features derived from the overlapper's in-overlap talk and the overlappee's pre-overlap talk. Using F0 mean as an example:

$$\texttt{erF0meanInRePre} = \texttt{erF0meanInReClear} - \frac{F0_{ee,pre} - \mu_{ee,clear}}{\sigma_{ee,clear}} \quad (6)$$

Here, $F0_{ee,pre}$ represents the mean F0 of the overlappee's pre-overlap talk, and the terms $\mu_{ee,clear}$ and $\sigma_{ee,clear}$ denote the mean and standard deviation of the overlappee's F0 computed in clear turns. In order to compete for the turn, overlappers might not modify the design of their incomings globally, relative to their talk in clear turns, but rather accommodate it locally to the overlappee's ongoing talk. Features computed within this context are intended to capture such potential modification.

The overlappee's in-overlap talk is also compared to the same talker's pre-overlap talk. Again, this is computed as a difference in z-scores between the overlappee's in-overlap and pre-overlap features:

$$\texttt{eeF0meanInRePre} = \texttt{eeF0meanInReClear} - \frac{F0_{ee,pre} - \mu_{ee,clear}}{\sigma_{ee,clear}} \quad (7)$$

Features computed over this context capture the potential modifications of prosodic features by overlappees com-

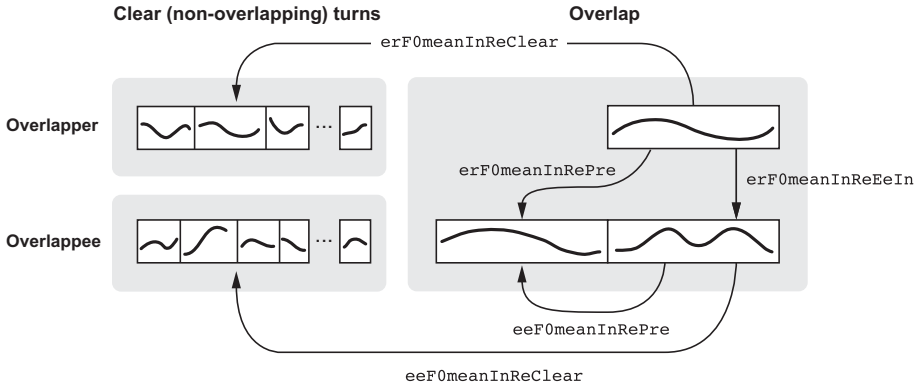

Fig. 2. Schematic illustration of overlap context types used in the analysis. Overlap contexts involving the overlapper begin with `er`, those involving the overlappee only begin with `ee`. The `F0mean` feature is used for illustration. For example, the feature denoted `erF0meanInReClear` compares the overlapper's mean F0 from the region of overlapping talk with the same overlapper's mean F0 measured from clear (non-overlapping) turns.

pared to their pre-overlap talk, in competitive and non-competitive overlaps.

Overlapper's in-overlap talk compared to overlappee's in-overlap talk (`InReEeIn`). This is computed as a difference between z-scores for features of the overlapper's and overlappee's in-overlap talk. For the mean F0 feature, for example, we have:

$$\texttt{erF0MeanInReEeIn} = \texttt{erF0meanInReClear} - \texttt{eeF0meanInReClear} \quad (8)$$

This context is included to capture the potential accommodation of prosodic features during overlap between the speakers. It refers to speakers' prosodic orientations to each other's talk in turn competition and overlap, as suggested by Szczepek-Reed (2006), and indicates whether speakers mark turn competition by reference to each other's prosodic modifications.

3.3.1.2. Units of time. F0, intensity and speech rate can be computed over units of speech of variable size. The choice of unit is important because it embodies the hypothesis that the selected unit is the one that participants orient to in their organisation and monitoring of ongoing talk, and management of turn competition in overlap. From our point of view the best analysis unit should, as Schegloff (2000), p. 15 puts it, "...offer itself not as an external analyst's imposition, but as an indigenous aspect of the participants' understanding of the organisation of overlapping talk".

Currently, no precise evaluation of different possible units and their relevance for turn competition in overlap exists. Overlappers could design only the beginning of their incoming in a prosodically distinct way to indicate turn competition early in the incoming. Alternatively, talkers may maintain a particular prosodic design until the resolution of the overlap. Also, overlappees may respond to the overlapper's incomings as soon as they realise there is overlap, or not at all, in which case the entire in-overlap portion would be a unit for a particular prosodic design. Given these considerations we evaluate several possible units within the overlap that may be designed in a prosodically distinctive way by overlappers and overlappees.

Specifically, we compute the overlappee's prosodic features over two units: the entire in-overlap speech and the in-overlap speech reduced by the 'reaction time', a time span during which overlappees potentially realise that the overlap is underway, so that they tune their response accordingly in the remainder of the overlap. We set the reaction time to be the duration of the overlapper's first syllable, since Schegloff (2000) suggests that it takes roughly the duration of a syllable for participants to notice that overlap is underway.

For the computation of overlappers' prosodic features we include units extending over K overlap initial words. The distribution of the number of overlapper's and overlappee's words in overlap has an interquartile range of $[1;5]$, therefore we vary K from 1 to 5. If $K = 5$ for example, we include all talk extending from the overlap onset up to 5 words as a unit. For many overlaps that contain less than 5 words, the entire in-overlap portion is included. Also we use pause delimited units (PDUs), where we include all speech preceding a pause of a minimum predefined length, which is varied from 0.1–1 s in 0.1 s steps. If there is no pause in the in-overlap section, the entire in-overlap section is taken as a unit. Finally, the overlapper's prosodic features are also computed over the entire in-overlap talk reduced by the reaction time (RT). In this case RT is defined as the duration of the overlappee's first two syllables in overlap.

3.3.2. Pausing

Pausing (PAU) features are given in Table 4. They indicate the frequency, position and length of pauses in both speakers' in-overlap talk. A pause is defined as a silence between words. Start and end times of each pause are available from the word-level alignments of the corpus. This

Table 4
Pausing features that describe the position, duration and frequency of pauses of the participants' in-overlap talk.

Feature name	Description
PAU features	
`erOnsetUponPause`	Indicates whether overlap onset is placed after a pause in the overlappee's turn (true/false)
`erPausePositionIncoming`	Position of the first pause in the overlapper's incoming (a value between 0 and 1 indicating the position relative to the number of in-overlap words)
`erPauseDurIncomingReClear`	Duration of the first pause in the overlapper's incoming relative to the mean and SD of the duration of that speaker's turn-internal pausing in clear segments (seconds)
`cmpPauseDurLastFirst`	Difference in duration between the overlapper's first pause in the incoming and the overlappee's last pause before overlap onset, normalised with respect to clear turns of each speaker (seconds)
`erPauseFreqReNrWords`	Number of pauses in the overlapper's incoming relative to the number of words in the incoming (numeric)
`erPauseDurInReOverlap`	Total duration of all pauses in the overlapper's incoming compared to the total duration of the overlap (seconds)
`erPauseDurInReClear`	Total duration of all pauses in the overlapper's incoming compared to the overlapper's mean pause duration in non-overlapped turns (seconds)
`eePauseFreqReNrWords`	Number of pauses in the overlappee's in-overlap talk relative to the number of words in overlap (numeric)
`eePauseDurInReOverlap`	Total duration of all pauses in the overlappee's incoming compared to the total duration of the overlap (seconds)
`eeOnsetPauseDur`	Duration of the overlappee's pause upon which the overlap onset takes place (seconds)
`eePauseDurInReClear`	Total duration of all pauses in overlappee's incoming compared to this overlapee's mean pause duration in non-overlapped turns (seconds)

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002

definition of pause indicates all silences in a speaker's signal, not only the ones that are perceivable as such by the participants. Early psychophysical studies suggest that the threshold for detection of an acoustic silence in conversational speech is close to 200 ms (Walker and Trimboli, 1982). More recently, Heldner (2011) has reported that the detection threshold for gaps in speaker changes is about 120 ms. However, automatically estimating a 'hearable' pause is a challenging problem and is not addressed here; for example, Heldner (2011) notes that gap detection thresholds vary substantially across individuals, and are influenced by factors such as musical training.

3.4. Overlap placement and completion features

These features relate to the placement of the overlap in time, and comprise three categories: duration, overlap onset position features and turn completion features.

Duration features are given in Table 5. Although durational properties of speech are often classed as prosodic features, our durational features are intended to capture the speaker's persistence in overlap, rather than (prosodic) variations in the length of individual syllables for example. The duration features express the entire duration of both the overlapper's and overlappee's in-overlap talk in terms of time, and also in a number of linguistic units such as syllables and words. The duration of the in-overlap talk is predicted to be closely associated with turn competition. As Jefferson (2003) notes, speakers sometimes compete for the turn by just keeping on talking in overlap. Competitive overlaps are thus expected to be longer events in which both speakers persist beyond the point of realisation that they are talking in overlap, and by doing so signal their interest in competing for the turn. Non-competitive overlaps are expected to be shorter and resolved soon after one of the participants realises that overlap is under way.

Overlap onset features are given in Table 6. These features characterise overlaps in terms of the place where the overlapper positions the overlap onset relative to the ongoing turn. These features are binary and include Jefferson (1983) overlap onset categories: `BlindSpot`, `Terminal`, and `Progressional`, as well as two further features, (`AtCompletion` and `SimStart`), that describe positioning of the overlap onset relative to the point of syntactic completion within the ongoing turn and the completion of the turn itself respectively.

Turn completion features are shown in Table 7. These features describe some remaining phenomena that are often found in overlap. Some of these features have previously been found to be related to overlap competitiveness and winning the turn (`Recycling`, `Completion`, `DelayedCompletion`) while for others no previous hypotheses exist. Turn completion features describe how both speakers, overlapper and overlappee, design their in-overlap turns, so like the prosodic features, the features have a value for both the overlapper and the overlappee.

Table 5
Durational features.

Feature name	Description
DUR features	
`eeBeforeDur`	The duration between the start of the overlappee's turn and the onset of the overlap, normalised by the duration of the overlappee's turn (seconds)
`inOverlapDur`	The duration of the in-overlap talk (seconds)
`erNrWordsInOverlap`	The duration of the overlapper's in-overlap talk (number of words)
`eeNrWordsInOverlap`	The duration of the overlappee's in-overlap talk (number of words)
`eeNrSyllablesInOverlap`	The duration of the overlapper's in-overlap talk (number of syllables)
`eeNrSyllablesInOverlap`	The duration of the overlappee's in-overlap talk (number of syllables)

Table 6
Overlap onset features. All of these features have a Boolean value.

Feature name	Description
ONSET features	
`AtCompletion`	Identifies all overlaps that start at any point of a possible syntactic completion within the ongoing turn. A stretch of speech is syntactically complete if it constitutes an interpretable clause within the conversational sequence in which it occurs. Elliptic clauses, answers to questions and response tokens are regarded as syntactically complete, e.g. Extract (4), l. 8.
`SimStart`	Overlaps in which participants start up simultaneously, e.g. Extract (4), ll. 5–6. A start up is defined as simultaneous if speakers are heard as starting at the same time, even though the precise timings are not identical.
`BlindSpot`	Overlaps in which the overlapper starts soon after the overlappee's turn begins, but cannot be counted as a turn incursion. Jefferson (1987) explains these as resulting from the delay in transition between speakership and listenership, so that overlappers who would start at the point of the current speaker's turn completion need 'a bit of space' before starting their turn during which the next turn is initiated, resulting in overlap close to the beginning of that turn.
`Terminal`	Overlaps located within the last phonological word of the turn. At this point the end of the turn is reliably projected, e.g. Extract (3), ll. 1–2.
`Progressional`	Overlaps in which a speaker starts upon a disfluency in the current speaker's turn. Pauses, filled pauses, stutters, repetitions or a combination of these are counted as disfluencies, e.g. Extract (3), l. 6.

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002

Table 7
Turn completion features. All of these features have a Boolean value.

Feature name	Description
COMPL features	
`Recycling`	Repetitions of two or more times of constituents of any length within a turn that contains overlap, where the overlap contains at least one repetition (e.g. the overlap in ll. 8–9 in Extract (5)).
`DelayedCompletion`	Overlaps where in-overlap speech is quitted and then continued, repeated or restarted after the overlap. Extract (4), l. 8.
`Completion`	Indicates whether the turn has been completed by the overlapper and overlappee.
`CutOff`	Indicates whether an overlapper or overlappee's turn that contains overlap is heard as ending abruptly, e.g. in a glottal closure.

3.5. Decision tree modelling

We use a decision tree classification paradigm (Breiman et al., 1984) to explain the relevance of the above features and their combinations as resources for turn competition in overlap. If these features serve to distinguish turn competitive incomings from non-competitive ones in the corpus, the decision tree model that makes use of these features will successfully classify overlap instances as competitive or non-competitive. An important advantage of the decision trees for our purposes is that the output tree is human readable and easily interpretable. The decision trees that are derived from the data can therefore suggest hypotheses on how turn-competitive and non-competitive overlaps are designed by participants using single features (prosodic and other) and their combinations. Decision trees were trained from our data using an implementation of Quinlan's (1994) C4.5 decision tree learner, as provided by the Weka toolkit.[4]

The success of classification is measured in terms of classification correctness, i.e. the percent of correctly classified overlap instances in the entire set of overlap instances. We compare the performance of the decision tree trained on different features and feature combinations to the performance of the *majority baseline* classifier, which classifies all instances as the class that occurs more often in the data. In our case, the majority class is the class of non-competitive overlaps and the correctness of the majority baseline is 51.68% for our data set. We use 10-fold stratified cross-validation for the evaluation of the decision trees. All results reported in this section are generated by repeating each 10-fold cross validation 10 times to minimize the effect of random variation in choosing the folds.

We first evaluate the contribution of each entire feature set to competitiveness classification. For this purpose we build an all-features decision tree and assess its performance relative to the majority baseline. If there is a significant improvement in performance over the baseline system, it means that at least some of the features from the set are useful for making competitiveness predictions. In this case we evaluate each feature separately to see which feature or features have contributed to the improved classification performance. It is expected that some features may have a strong contribution when used alone, while others may be more useful when combined with other features from the same set. For this reason, for each feature, we train two decision trees: one using that feature only (*leave-one-in*) and another using all other features but that feature (*leave-one-out*). We compare the performance of each of these decision trees to the performance of the majority baseline classifier, as well as to that of the tree trained using all features. An increase or decrease in correctness indicates the relevance of each single feature either as a single resource or in combination with other features. In the final step of the analysis, the best-performing decision tree is inspected to establish how prosodic and overlap placement features are combined to arrive at competitiveness decisions in the classification.

[4] Available from `http://www.cs.waikato.ac.nz/ml/weka/`.

4. Results

To identify the prosodic and overlap placement features that characterise turn competition, we first assess the utility of the prosodic and overlap placement feature sets as individual turn competitive resources, and then describe the potential interactions between these feature groups. According to the Shapiro–Wilk test, the null hypotheses that the data follow a normal distribution could be retained for all result sets. Consequently, in the following, all significant values are reported as indicated by a two-tailed paired t-test, $p < 0.05$.

4.1. Prosodic resources for turn competition

Table 8 shows the performance of each of the prosodic feature sets and their combination in the classification of

Table 8
Classification correctness of the prosodic decision tree models of the overlapper's and overlappee's in-overlap speech trained on the features shown in the first column of the table in leave-one-in and leave-one-out mode. Statistically significant changes compared to the all-features decision tree (first row) are marked with an asterisk.

Feature set	Leave-one-in	Leave-one-out
ALL	65.07	
F0	62.70*	64.67
INT	64.94	62.45
SR	58.75*	65.33
PAU	60.01*	65.33
F0-INT	65.15	58.99*
F0-SR	62.93	65.00
F0-PAU	62.06*	64.83

overlaps according to competitiveness. When used alone, each prosodic feature set significantly outperforms the majority baseline (51.68%), indicating that each of the prosodic feature classes is used by participants individually as a turn competitive resource. However, these individual resources differ in their robustness. The best predictors are F0 and intensity, whereas the lowest classification scores are achieved by speech rate features.

The combination of all prosodic features (ALL classifier) gives better classification performance than each of the individual feature sets alone. The performance of the ALL classifier is significantly higher than that of the pausing, speech rate and F0 feature sets and moderately higher than that of intensity features. This means that, generally, combining prosodic features contributes towards the classification more than using the feature sets individually.

The combination of F0 and intensity (F0-INT) is the strongest predictor of turn competitiveness of an overlap. The performance of the F0-INT classifier is close to that of the ALL classifier. Removing both F0 and intensity features from the ALL classifier results in a significant degradation of predictive performance, which indicates that speech rate (SR) and pausing features are less important than F0 and intensity.

From these results we conclude that prosodic features are associated with turn competition. However, our results reveal that it is prosodic feature clusters rather than individual prosodic feature sets which are the best predictors of turn competition. The main such prosodic cluster is the combination of F0 and intensity.

4.2. Overlap placement and completion resources for turn competition

Table 9 shows the classification correctness results of the overlap placement feature sets. Again, all classifiers gave a statistically significant improvement over the majority baseline (51.68%). The best predictors of turn competitiveness of the overlaps are turn completion (COMPL) features. The classifier trained on COMPL features significantly outperforms classifiers trained on durational and overlap onset features, whose performance is approximately the same.

Table 9
Classification correctness of the overlap placement/completion feature decision tree models of the overlapper's and overlappee's in-overlap speech, trained on the features shown in the first column of the table in leave-one-in and leave-one-out mode. Statistically significant changes compared to the all-features decision tree (first row) are marked with an asterisk.

Feature set	Leave-one-in	Leave-one-out
ALL	74.05	
DUR	65.99*	72.21
ONSET	66.73*	74.50
COMPL	71.01*	69.23*

Removing turn completion features results in a significant degradation of performance compared to the ALL classifier, while there is a degree of redundancy between durational and overlap onset features, which can be removed from the ALL classifier without a significant loss in performance. Nevertheless, all of the overlap placement feature sets contain some features used for competitiveness classification, since combining all features together in the ALL classifier performs significantly better than each of the three feature sets alone.

The decision tree in Fig. 3 shows which particular features from the three feature sets are the most relevant in the competitiveness classification and how they combine in the ALL classifier. The feature `Terminal`, i.e. the positioning of the overlap onset at the last phonological word of the ongoing turn, close to its completion and the TRP, best discriminates between competitive and non-competitive overlaps. Terminal overlaps are generally non-competitive unless overlappers use recycling, which can indicate turn competition even from the terminal overlap onset position. In non-terminal overlaps, the use of recycling by either overlapper or overlappee can be indicative of turn competition. If there is no recycling present, the overlappee's sudden termination of the ongoing turn indicates that the incoming is classified as turn competitive.

Non-terminal overlaps longer than three of the overlapper's words are generally competitive, regardless of their other positioning within the turn. This means that even the non-incursive overlaps starting earlier in the ongoing turn and close to the TRP, like simultaneous starts and blind spots, can develop into turn competitive events, and that this will be indicated by the participant's persistence in overlap. Also, it should be noted that the decision tree selects the durational features expressed in terms of number of words rather than time, which suggests that participants monitor the ongoing speech by orienting to linguistic units, a hypothesis that could be investigated in perceptual experiments.

Overlaps shorter than three words are competitive if they have a delayed completion, and non-competitive if they are simultaneous starts. For the remaining overlaps, the overlapper's drop-out after an overlap of three syllables or shorter is found in non-competitive overlaps, while persisting beyond this duration is still turn competitive. Schegloff (2000) proposed that overlap is managed on a beat-by-beat basis, where a beat roughly corresponds to a syllable. According to him, participants can develop turn competition by persisting in overlap beyond two syllables and modifying the prosodic features of talk on each beat. Our results on this final group of overlaps seem to support this idea of very short units being sufficient for the participants to organise overlap. However, in order to directly assess Schegloff's claim, it is necessary to combine prosodic and

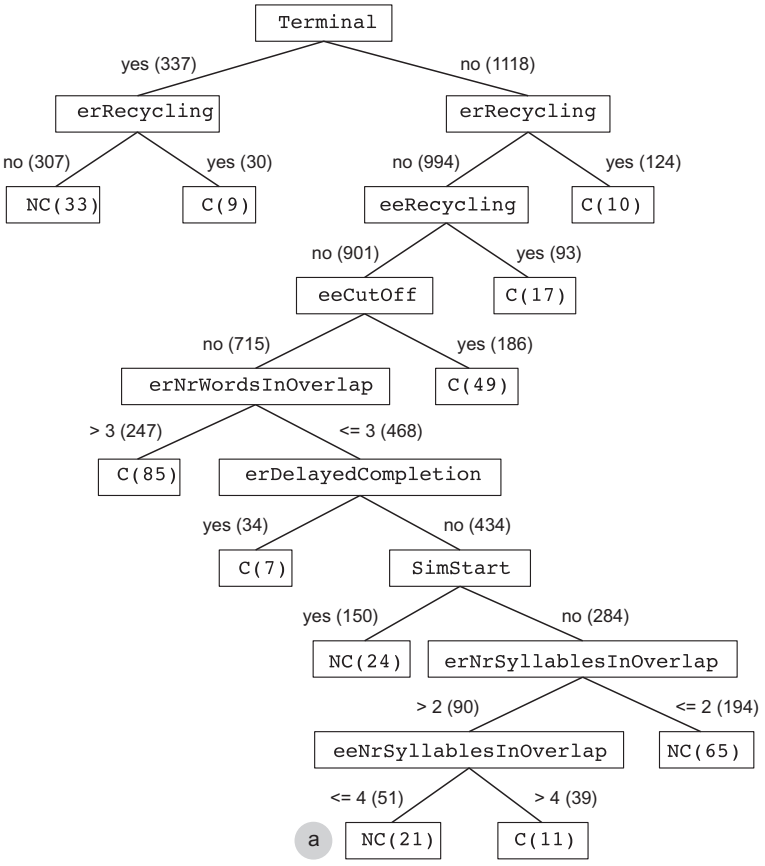

Fig. 3. A model of turn competition in overlap constructed using overlap placement features only. The numbers in brackets on the branches indicate the number of overlap instances going down the branches of the decision tree. Leaf nodes labelled with C indicate a competitive decision, those labelled with NC indicate a noncompetitive decision. The numbers in brackets on the leaf nodes indicate the number of misclassified instances. Other nodes are labelled with the name of an overlap placement feature as given in Tables 5–7.

overlap placement features in an integrated decision tree model (see next section).

From these results we hypothesise that the overlap placement features we investigated are relevant turn competitive resources, and that participants use combinations of these resources to indicate turn competition in overlap.

Our results do not support the claim that overlaps positioned at or around the TRP (i.e. simultaneous starts, terminal and blind spot overlaps in our feature set) are inherently non-competitive, while turn incursive overlaps are competitive. Rather, the results indicate that both competitive and non-competitive overlaps can occur at a range of different places in the turn. Non-competitive overlaps are possible in the middle of the turn, in which case they are mostly short. Conversely, a turn competitive overlap can develop from an overlap onset positioned around the TRP if participants employ further turn competitive devices. For example, the results suggest that short simultaneous starts are non-competitive, as participants resolve the overlap upon the realisation that it is underway, while longer simultaneous starts are turn competitive since participants persist in overlap, indicating their interest in the turn. Likewise, recycling can turn a terminal overlap into a competitive one. Overlappers' recycling of turn beginnings in terminal position can be understood as an overlap absorbing strategy (Jefferson, 2003; Schegloff, 1987). The suggestion is that overlappers assume that turn beginnings will not be heard, as they are uttered in overlap with the terminal token of the overlappee's turn. In this case recycling may be used to ensure that the in-overlap turn beginning is heard, which provides the necessary basis for further continuation of the turn.

4.3. Integrated prosodic and overlap placement model of turn competition

The performance of a decision tree classifier that utilises both prosodic and overlap placement/completion features is shown in Table 10.

Table 10
Decision tree performance (% correct) for overlapper's and overlappee's combined prosodic and overlap placement/completion features. All classifiers gave a statistically significant improvement over the majority baseline (51.68%). Statistically significant changes compared to the all-features decision tree (first row of the table) are marked with an asterisk.

Feature set	Leave-one-in	Leave-one-out
ALL	74.18	
PROSODIC	64.70*	74.14
OVERLAP PLACEMENT	74.14	64.70*

The combined model composed of both prosodic and overlap placement features outperforms the classifiers built using each set of features individually. This indicates that a combination of features from both sets is more strongly predictive of overlap competitiveness than prosodic or overlap placement features alone. In this combination, however, the overlap placement features are significantly stronger predictors than the prosodic ones.

The decision tree in Fig. 4 shows how prosodic, overlap placement and turn completion features combine in the integrated model of turn competition. The top part of the tree is equivalent to the decision tree built from overlap placement features (Fig. 3), indicating that participants' recycling is the main turn competitive resource in both terminal and non-terminal positions.

The prosodic features are only relevant for turn competition in non-terminal overlaps, in which neither speaker employs recycling. The overlapper's modification of the intensity range compared to his norm (i.e., talk in clear turns), which takes place upon realisation of in-overlap talk by the overlappee, is the feature at the top of this subtree. If the intensity range is narrower than the norm, the overlapper needs to prolong the in-overlap talk past three words, in order to signal competition. A narrow intensity range in combination with the overlapper's quitting overlap after three words or less is a feature of non-competitive overlaps.

The tree model furthermore shows how prosodic features interact with overlap placement features, in overlaps in which the intensity range is widened. The tree built from overlap placement features indicates that simultaneous starts are competitive if participants persist in overlap. The combined model shows that persisting in overlap goes hand-in-hand with an increase in the overlapper's mean F0 compared to the norm for the clear turns. If no such increase takes place, even longer simultaneous starts are more likely to be non-competitive.

The competitiveness of overlaps whose onsets are placed in the middle of the ongoing turns, but which do not have a delayed completion, is also realised by prosodic modification. In very short overlaps, where the overlapper's talk in overlap is shorter than 3 syllables, widening the F0 range above the norm can be used to compete for the turn, while non-competitive overlaps usually have a narrower F0 range. This result suggests that prosodic modifications performed over short units like syllables are indeed employed

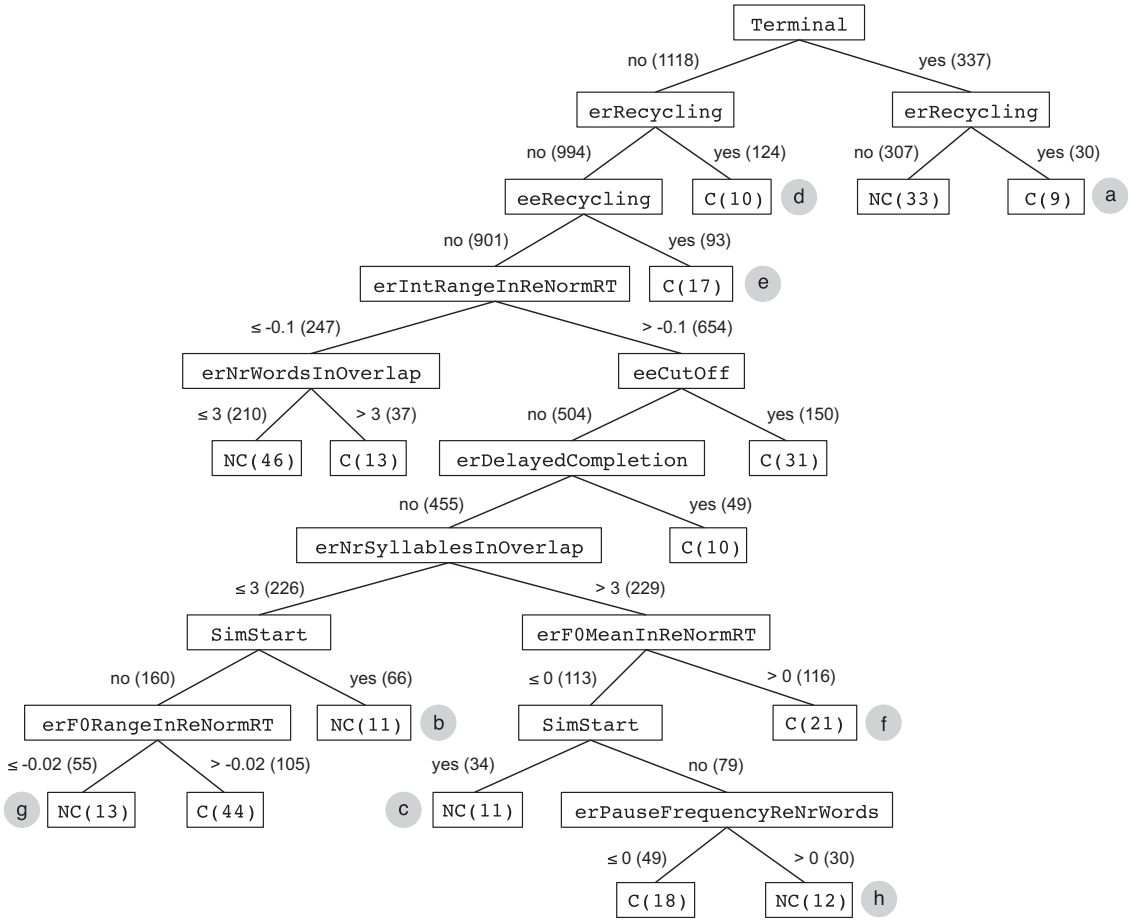

Fig. 4. The combined model of turn competition in overlap, including all features (i.e. prosodic, overlap onset and turn completion features). The numbers in brackets on the branches indicate the number of overlap instances going down the branches of the decision tree. Leaf nodes labelled with C indicate a competitive decision, those labelled with NC indicate a noncompetitive decision. The numbers in brackets on the leaf nodes indicate the number of misclassified instances. Other nodes are labelled with the name of a feature as given in Tables 3–7.

for turn competition in very short overlaps. This supports Schegloff's (2000) claim, discussed above, that overlap management takes place on a syllable-by-syllable basis, and the results from our integrated model show that prosodic modification over each syllable indeed plays a role in the realisation of turn competition.

Finally, in turn medial overlaps longer than 3 syllables, it is absence of pausing that characterises competitive overlaps. In non-competitive turn-medial overlaps, overlappers may pause to monitor the ongoing turn and give the overlappee the opportunity to approach the TRP before the overlapper restarts with his turn. A pause is equivalent to temporarily relinquishing overlap upon realisation that it is underway. However, in turn competitive overlaps, it appears that overlappers demonstrate no such orientation towards the overlappee completing his turn.

5. Discussion

In this study, we used decision tree analysis of a large corpus of conversational speech to investigate the resources that participants might employ and orient to when competing for the speaking turn. A wide range of features were extracted from the corpus of overlap instances, including both prosodic features (e.g. F0, intensity, speech rate) and those related to the placement of overlapping talk (duration, the position of overlap onset in the current speaker's turn, and other phenomena associated with overlap such as recycling). Decision tree models were trained on these features, and tested on their ability to discriminate competitive overlaps from non-competitive overlaps. Furthermore, decision trees were trained on subgroups of the available features (e.g., prosodic features only) in order to assess the contribution of that subgroup to overlap classification performance.

The decision tree models that derive from these analyses shed light on how competitive and non-competitive overlaps differ. While these models do not constitute direct evidence of the feature clusters actually used by participants in conversation (for example, participants probably employ measurements of F0 and duration that are perceptually scaled, rather than in the linear units used here), our study provides empirically based, testable hypotheses about human behaviour. These can be investigated in future perceptual studies, for example in the way that Hjalmarsson (2011) has tested out perceptually some hypotheses about cues to turn-finality that derive from earlier corpus-based research.

More specifically, our decision tree models gives rise to the following hypotheses:

1. Turn competition can be initiated by a new speaker at different points in the current speaker's turn; likewise, non-competitive overlaps can occur at different points in the turn;
2. Turn competition in overlap that is initiated around points of possible turn completion is realised using different resources than when initiated turn-medially;
3. The combination of F0 and intensity is the most prominently used prosodic resource for turn competition;
4. Positional features of turn-design, notably recycling, play a major role in indicating turn competition.

Each of these hypotheses will now be discussed, and illustrated by reference to specific instances of overlap drawn from our corpus. By doing so, we aim to show how the general – and therefore necessarily abstract – properties of overlap and turn competition identified in our model may be realised on particular occasions in situated talk-in-interaction.

1. Turn competition can be initiated by a new speaker at different points in the current speaker's turn; likewise, non-competitive overlaps can occur at different points in the turn

Our integrated model (Fig. 4) suggests that it is possible to compete from a wide range of different places in the ongoing turn. Our model therefore does not substantiate the assumption commonly made in previous work that turn competitive overlaps have to be placed earlier in the turn, while overlaps at TRPs are not competitive (French and Local, 1983; Wells and McFarlane, 1998; Schegloff, 2000). Rather, our model suggests that incoming in overlap is a resource for action that participants have at their disposal at any time. We expected this to be the case for non-terminal competitive overlaps, as was illustrated in Extracts (1)–(3) in Section 3.2. It was less expected that an overlap in *terminal* position may also be competitive, as in Extract (5) below (which is a continuation of Extract (3)):

(5) ICSI_Bmr007_111:

```
8   fl6: [so] the question i:s
         >*you know*< (.)
         are there really more
         overlaps happening .hhh
         (0.9)
         than there would be in a
         two-person (0.2)*[party]* and
9 > ml3:      .hhh         [let] (.)
         [let m let me rephrase what
                              I'm saying]
10 fl6: [and there well may be *but*]
11 ml3: cuz I don't think I'm getting
         it across what what I what
         (0.5) I shouldn't use words
         like ''nice'' because
```

In line 9 the onset of *m13*'s first "let", in overlap with *f16*'s "party", is in terminal position, because it is within a TRP (Wells and McFarlane, 1998). *f16* has projected the end of her turn through the pitch accent on "two-person": the main pitch and loudness prominence is on "two", which thus marks the start of the TRP. This can be seen in Fig. 5.

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002

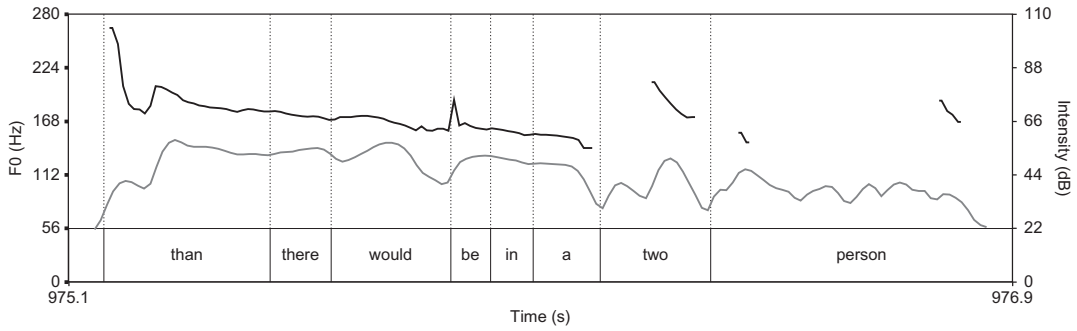

Fig. 5. Fundamental frequency (black) and intensity (grey) contours for line 8 of Extract (5), showing female speaker *f16*'s pitch accent on "two-person", with the main pitch and loudness prominence on the word "two".

In this extract, *m13* recycles "let" twice, eventually securing the floor. The decision tree in Fig. 4 shows that of the 337 instances of terminal overlap, 30 were classified as competitive (see (a)) and were characterised by recycling of this type. In 21 cases (including the example in Extract (5) above) the classification agreed with the human annotation.

Turning to non-competitive cases, in both terminal and non-terminal positions some overlaps may occur as by-products of the latitude that exists in the basic system of turn-taking organisation. Terminal non-competitive overlaps arise from the fact that the TRP can start slightly before the end of the current speaker's turn, e.g. Extract (3), lines 1–2. (cf. Couper-Kuhlen, 1993; Wells and McFarlane, 1998). One type of non-terminal non-competitive overlap is the simultaneous start up, which is a by-product of self selection by more than one party, one of whom may be the current speaker selecting to continue having reached a TRP (e.g., the overlaps in lines 3–4 and 5–6 of Extract (4)). Fig. 4 shows that 100 such non-competitive simultaneous starts appear in the model (see (b) and (c) in the figure). While both these types of non-competitive overlap may be viewed as occurring by accident rather than design, there are also various types of overlap that are designed to be non-competitive. These include response tokens, choral overlaps such as greetings and toasts, as well as collaborative completions done in overlap. Although those subtypes were not included in the collection of overlaps here, a substantial number of *other* non-terminal, non-competitive overlaps that are not simultaneous starts appear in the decision trees (51 in the tree constructed from overlap placement features only, node (a) of Fig. 3; 85 in the combined model, nodes (g) and (h) of Fig. 4).

Given that an incoming speaker can choose to position both competitive and non-competitive overlapping incomings at a range of different points in the ongoing turn, it is clearly necessary for listeners, and particularly the current speaker, to be able to differentiate between them, and to respond accordingly. How this is achieved in our model will now be discussed.

2. Turn competition around points of possible turn completion is realised using a different range of resources than turn-medially

We predicted that the location of the overlap onset within the ongoing turn would have a bearing on the design of the incoming. The rationale behind this previously unaddressed hypothesis is that different resources may be required to realise competition in an environment where competition might appear unnecessary (i.e. at a TRP), than it would when the incoming clearly violates the current speaker's right to the turn.

In our model (Fig. 4), turn competition around points of possible turn completion is realised by different means than when initiated turn medially. This is evidenced by the difference in designs of overlapping incomings in terminal and non-terminal positions: according to the model, prosodic features (of intensity, F0, pause) may be used to mark turn competition in non-terminal position, but not in terminal position. In terminal position, avoiding recycling may be sufficient for an overlap to be treated as non-competitive, since according to our model, terminal competitive incomings are characterised by recycling by the overlapper (see (a) on Fig. 4). This is evident in the example of the recycling of "let", in line 9 of Extract (5), discussed above.

Our model suggests that recycling may also occur in non-terminal competitive incomings, in which case it may again be the overlapper who recycles: see (d) on Fig. 4 ($n = 124$). Alternatively, it may be the overlappee who does the recycling, see (e) on Fig. 4 ($n = 93$) as in the following example (Extract (6)) from our corpus. Here, *m13* is the overlappee in the face of a competitive incoming from *f16*. He recycles "I'm saying if I -" , and *f16* drops out:

(6) ICSI_Bmr007_113:

```
1     m13: I was commenting about this huh
           huh huh
2     f16: OK
3     m13: I'm saying
4  >  f16: [all I'm saying is that from the]
5     m13: [if I (0.2) I'm saying if I have this]
           complicated thing in front of me
```

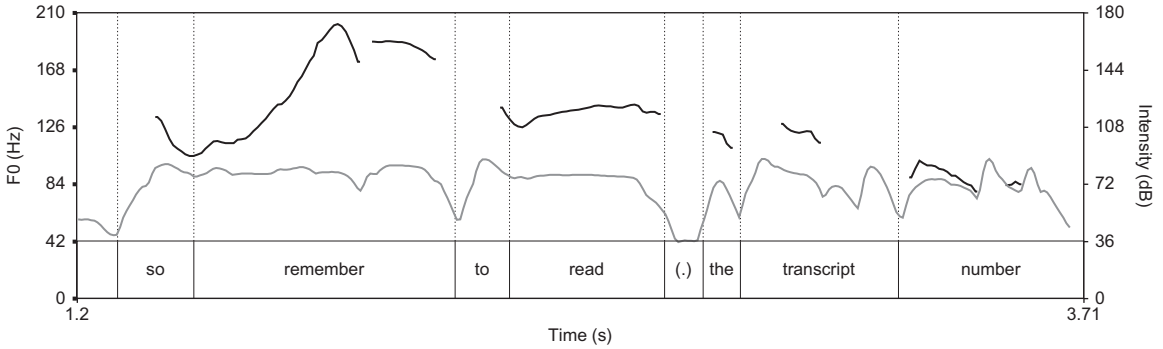

Fig. 6. Fundamental frequency (black) and intensity (grey) contours for line 6 of Extract (7), spoken by male speaker *m11*.

However, turn-medial competitive incomings may alternatively occur without recycling, as in Extracts (1)–(3) in Section 3.2. In those cases there was no recycling by either party, but the incomer raised pitch and/or loudness relative to her norm (see (3) below for further discussion). Thus, in non-terminal position, in the absence of recycling, a modification of prosodic features within the overlap may be needed to display the difference between a competitive and a non-competitive overlap. Different turn designs may need to be used depending on where in the ongoing turn the overlap starts. This suggests that listeners closely monitor the ongoing talk in real time and that speakers deploy phonetic and temporal resources accordingly in order to accomplish actions such as competing for the floor.

3. The combination of F0 and intensity is the most prominently used prosodic resource for turn competition

Decision tree classifiers trained on prosodic features outperformed a majority baseline classifier, suggesting that prosodic features may be used by participants to compete for the turn. This finding is consistent with previous work on the role of individual prosodic features, which suggests that intensity (Lee et al., 2008), F0 (Kurtić et al., 2009), speech rhythm (Couper-Kuhlen, 1993) and speech rate (Kurtić et al., 2010) can be employed by participants to signal turn competition.

However, our results suggest that combinations of features, rather than single features, may be functionally most important. The main such prosodic cluster is the combination of F0 and intensity. This is consistent with the finding of French and Local (1983) that the combination of pitch and loudness is the main resource used (and oriented to) by the participants in turn competition. Specifically, they suggest that overlapping talkers compete for the turn by raising their pitch and loudness.

An example of this is found in lines 5–6 of Extract (1), reproduced here as Extract (7):

(7) ICSI_Bmr018_566:

```
5   fl6: so and maybe we won't
         [laugh this time also]
6 > ml1: [so    remember    ]
         to read (.) the transcript
         number (0.2) so that uh everyone
         knows that what it is
```

The corresponding F0 contour for the overlapper (male speaker *m11*) is shown in Fig. 6. Here, *m11* raises pitch and loudness. His "so remember" comes in at a peak intensity of 84 dB and his F0 starts at 129 Hz, rising to 201 Hz. For out-of-overlap talk, speaker *m11* has a pitch range of 84–223 Hz, with a mean of 108 Hz; his mean intensity is 76 dB, with an intensity range of 48–85 dB. So he appears to be "high and loud" relative to his norm.[5] The decision tree of Fig. 4 correctly classifies this example as a competitive overlap; traversing the tree from top to bottom, the right-hand branches are taken after the nodes labelled `erIntRangeInReNormRT`, `erNrSyllablesInOverlap` and `erF0MeanInReNormRT`, terminating at a competitive decision marked by (f).

4. Positional features of turn-design, notably recycling, play a major role in indicating turn competition

According to our model (Fig. 4), the presence of recycling is the single most important feature of competitive overlaps. The role of recycling in the resolution of overlaps has been described in conversation analytic studies by Jefferson (2003) and Schegloff (1987, 2000), who point out that recycling of lexical material by one speaker serves to 'absorb' the overlapping talk being produced by the other speaker. The recycling speaker is effectively putting the progression of his own turn on hold, until the overlapping speaker drops out. As the model in Fig. 4 shows, the recycler may be the overlapper who uses recycling to sustain his bid for the floor, exemplified under (1) above. Alternatively, it may be the overlappee who recycles, in a bid to fend off the turn-competitive incoming from the overlapper, exemplified under (2) above.

In either case, the other participant in the overlap needs to be able to recognise in real time that the speaker is indeed recycling. Recycling is usually characterised, as it was for the purposes of this study, in terms of the repetition of 'constituents', e.g. of a syllable, a word, a phrase or part of a phrase recognisable by the repetition of the sequence of consonants and vowels that make up

[5] Note that the dB figures reported here are based on Praat's default sound pressure calibration (Boersma, 2001), because the ICSI corpus does not include recordings of a calibration sound with a known sound pressure level.

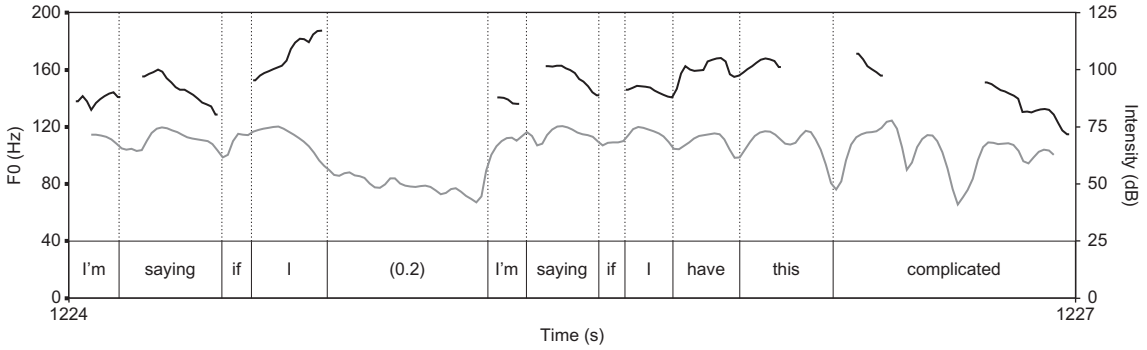

Fig. 7. Fundamental frequency (black) and intensity (grey) contours for lines 3 and 5 of Extract 6 uttered by male speaker *m13*.

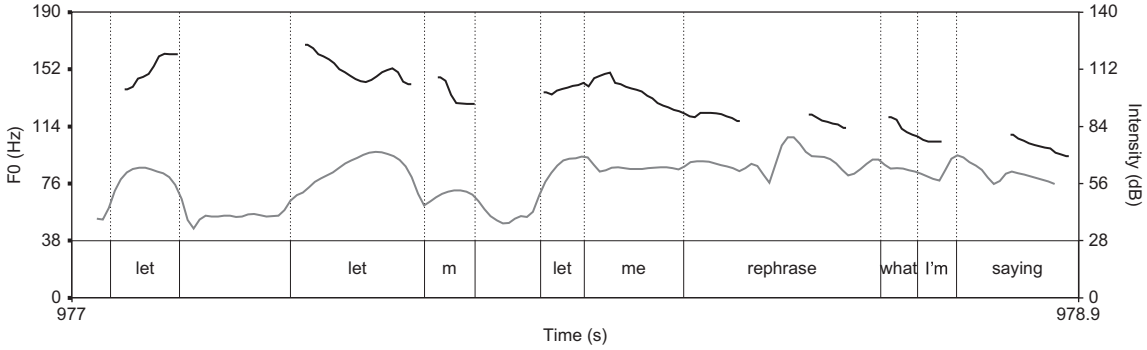

Fig. 8. Fundamental frequency (black) and intensity (grey) contours for line 9 of Extract 5 uttered by male speaker *m13*.

the constituent. Additionally, however, it appears that recycles in overlap have prosodic characteristics. There is a prosodic relationship between the original material and its recycle: the chunks seem to stay around the same relatively high region of the speaker's pitch range and may have a similar loudness. This is evident in the example (Extract (6)) discussed under (2) above, as shown in Fig. 7.

Speaker *m13* is the overlappee in the face of a competitive incoming from *f16*. He recycles "I'm saying if I -", and *f16* drops out. The F0 and intensity contours of the two versions of "I'm saying (if)" are very similar, and the F0 contours of "saying" peak at almost the same value (160 Hz vs. 162 Hz).

Another example of recycling, this time by an overlapper, was presented under (1) above as Extract (5). Speaker *m13* recycles "let" twice, eventually securing the floor. For out-of-overlap talk, speaker *m13* has a pitch range of 84–223 Hz, with a mean of 108 Hz. It can be seen in Fig. 8 that the three productions of "let" fluctuate in pitch direction, in the upper part of the speaker's range, with a peak of 168 Hz, before descending quite rapidly from approximately 145 Hz on "me" to 95 Hz on "saying". The intensity on each syllable is also relatively high. These prosodic features may help the listener to recognise in real time that a string of syllables is indeed a recycle. This could be especially valuable where the lexical material may be hard to decode, for example if there is just one syllable, as in "let".

Although the prosodic properties of recycles have not been systematically analysed in this study, examples such as these suggest that they may share an interesting feature with non-recycled competitive incomings such as the "so remember ..." example in Extract (7), namely the sustention of F0 (and possibly intensity) at a relatively high level, until the overlap is at or close to the point of resolution. If future research bears this out, it will mean that the role of prosodic features in the management of turn competition is more pervasive than the model presented in Fig. 4 would suggest.

6. Conclusion

Researchers interested in overlapping talk, irrespective of disciplinary background, have recognised that there is a fundamental distinction between accidental overlap and deliberate overlap. The mechanisms that underlie accidental overlap have been of particular interest to researchers designing speech-based computer systems that interact with human speakers; for instance, some researchers have endeavoured to indentify properties of the turn in progress that might predict whether the next speaker might start in overlap (e.g. Gravano and Hirschberg, 2011). It has been argued, e.g. by Heldner and Edlund (2010), that the fre-

quency of such accidental overlaps is counter-evidence to a model of turn-taking, attributed to researchers in Conversation Analysis, that assumes that conversation participants time their incomings with fine precision and that overlaps should therefore occur rarely if at all. In fact, since the publication forty years ago of the original and much-cited paper on turn-taking organisation by Sacks et al. (1974), those researchers and their followers have illustrated the range of types of overlap that frequently occur and have demonstrated the precision with which overlaps are initiated (e.g. Jefferson, 1983) and resolved (e.g. Schegloff, 2000).

While fully acknowledging the frequency of overlaps, and also the possibility that this may vary quite greatly across cultures (cf. Sidnell, 2001) and across speech activities, Conversation Analysis researchers such as Schegloff (2000) have held to the principle of 'one party at a time' as a fundamental design feature of turn-taking organisation. A simplistic version of this principle is embodied in some speech dialogue systems:

> "*Currently, the most common method for determining when the user is willing to yield the floor consists in waiting for a silence longer than a prespecified threshold* ..." Gravano and Hirschberg (2011, p. 601)

As these authors and others (e.g. Heldner and Edlund, 2010) point out, such a simple system is doomed as it fails to take account of turn final overlaps as well as turn medial silences. Yet it bears witness to the intuitive sense of the principle of "one party at a time", which all researchers seem to accept at some level. Specifically, it is accepted that while some overlaps may be accidental, others are used to deliberately compete for the floor. For example, the overlap scheme of Gravano and Hirschberg (2011) includes 'interruption' and 'butting-in', although they exclude these types from their own analysis (see Introduction above). The fact that turn-competition is an option for participants is itself evidence for the fundamental principle that at a given moment one party has the right to the floor.

In this study we have shown that much of the systematicity of the acoustic and positional properties of overlap can be revealed by using methods that explicitly acknowledge that speakers use overlap for interactional purposes, notably to compete for the turn, as well as for non-competitive purposes. Overlaps, particularly those we have described as competitive, may appear to the casual observer to be chaotic, disrupting what is assumed otherwise to be the smooth flow of conversation. However, our results, and the earlier research on which they build, indicate that in fact these overlaps are systematic in their design and organisation. These findings could be therefore incorporated into speech dialogue systems should it be deemed desirable (Reidsma et al., 2011).

Incorporating speakers' social actions, such as competing for the turn, into the analysis of overlapping talk gives rise to its own methodological challenges, which the current study has begun to address but does not claim to have completely solved. The broad notion of 'turn-competition' appears to be robust, widely accepted by researchers and explicitly oriented to by conversational participants. However, this does not exclude the possibility that there are subtypes of competitive incoming that may be designed in different ways. The case of non-competitive incomings is still more challenging in this respect. To 'not compete' for the turn yet to do so by talking in overlap is paradoxical: if a participant does not want the floor, then there is no *prima facie* reason to speak at all. The implication is that non-competitive incomings may include a range of different social actions. Some have already been described elsewhere but excluded from the present analysis: these include choral overlaps, collaborative completions and continuers/response tokens, each of which occurs at a sequentially distinct place in conversation, and yet they have in common that the incoming speaker, rather than competing for the floor is aligning with the action in progress in the overlappee's talk and/or affiliating with the overlappee's stance (Stivers, 2008). More fine-grained interactional analysis of the large class of 'other' non-competitive incomings from the current study is likely to reveal further subtypes, some of which may also be characterised as aligning or affiliative. Such subtypes may turn out to be characterised by distinct designs, in terms of prosodic and positional features. For example, the overlapper may use a pitch contour that matches the contour just used by the previous speaker/overlappee. This design was found by Gorisch et al. (2012) when investigating short turns, very often produced in overlap, that align with the talk produced by the preceding speaker.

When managing competitive and non-competitive overlapping talk, conversational participants closely monitor others' speech in real time. According to our model, various features have to be tracked in this way, in conversations conducted in American English, for example the pitch of the speaker's voice from the start of his turn relative to his habitual range. One may then wonder about the generality of such features. Are the positional and phonetic characteristics of overlapping talk similar across languages? If there are differences, do they relate to properties of the specific accentual/intonation systems of the language? Do phonetically-defined subtypes of overlapping talk have the same interactional functions in different languages? Investigating such questions should throw light on issues related to second language learning and intercultural communication. It may also contribute to more theoretical debates about the ontogeny and evolution of spoken communication: cross-linguistic comparison will shed light on how multi-speaker simultaneity is handled in different linguistic systems, deepening our understanding of the role of language in interaction and the mechanisms humans have for handling simultaneity in an interaction system designed for sequential turn-taking. This will contribute to the specification of the 'human interaction engine', described by Levinson (2006). We suggest that future

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002

research of this kind will benefit from an interdisciplinary approach that combines computational and speech science methods with an interactional phonetic perspective that is informed by Conversation Analysis.

Acknowledgements

The research reported here was supported by a University of Sheffield Project Studentship. Preparation of the article was facilitated by UK Arts and Humanities Research Council Grant 1-62874195. We are grateful to our annotators for their time and effort; to Ahmed Aker for invaluable assistance at various stages of the research; to Gareth Walker and John Local for their sustained interest and encouragement; and to Jens Edlund and an anonymous reviewer for their constructive comments on an earlier draft.

Appendix A. Transcription conventions

The transcription conventions are an adaptation of the usual format employed in the conversation analysis literature (Jefferson, 2004, p. 2). The conventions are as follows:

Overlapping talk:

`[` Opening square brackets aligned across adjacent lines denote the onset of overlapping talk.
`]` Closing square brackets indicate where the overlap ends.

Pauses:

`(.)` A pause of less than 0.2 seconds.
`(0.5)`, `(1.15)` Timed pauses of 0.5s and 1.15s respectively.

Speech tempo:

`> <` inward arrows denote faster speech.
`< >` outward arrows denote slower speech.

Other symbols:

`.hhh` Denotes an inbreath (note the preceding full stop).
`hhh` Denotes an outbreath.
`* *` The talk in between the asterisks is uttered in creaky voice.
`>` Denotes a significant line of interest, discussed in the text.

References

Adda-Decker, M., Barras, C., Adda, G., Paroubek, P., de Mareüil, P.B., Habert, B., 2008. Annotation and analysis of overlapping speech in political interviews. Proc. 6th Internat. Language Resources and Evaluation Conf. (LREC'08). Marakech, Morocco.

Barkhuysen, P., Krahmer, E., Swerts, M., 2008. The interplay between auditory and visual cues for end-of-utterance detection. J. Acoust. Soc. Amer. 123, 354–365.

Bavelas, J., Coates, J., Johnson, T., 2002. Listener responses as a collaborative process: The role of gaze. J. Comm. 52, 566–580.

Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot International 5, 9/10, 341–345.

Breiman, L., Friedman, J., Olshen, J., Stone, O., 1984. Classification and Regression Trees. Wadsworth, Belmont, CA.

Carletta, J., 2007. Unleashing the killer corpus: Experiences in creating the multi-everything AMI meeting corpus. Lang. Resour. Eval. 41, 181–190.

Cetin, O., Shriberg, E., 2006. Overlaps in meetings: ASR effects and analysis by dialogue factors, speakers, and collection site. In: Proc. 3rd Joint Workshop on Multimodal Interaction in Related Machine Learning Algorithms (MLMI'06), Washington, D.C.

Couper-Kuhlen, E., 1993. English Speech Rhythm: Form and Function in Everyday Verbal Interaction. John Benjamins, Amsterdam/Philadelphia.

Dellwo, V., Ferragne, E., Pellegrino, F., 2006. The perception of intended speech rate in English, French, and German by French speakers. Proc. Speech Prosody 2006. Dresden, Germany.

Dhillon, R., Bhagat, S.H.C., Shriberg, E., 2004. Meeting recorder project: Dialog act labelling guide. Technical Report TR-04-002. International Computer Science Institute (ICSI).

French, P., Local, J., 1983. Turn-competitive incomings. J. Pragmatics 7, 701–715.

Gardner, R., 2001. When listeners talk: Response tokens and listener stance. In: Pragmatics and Beyond New Series, vol. 92. John Benjamins Publishing Company, Amsterdam.

Goodwin, C., 1980. Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. Sociol. Inquiry 50, 272–302.

Goodwin, M., Goodwin, C., 1986. Gesture and coparticipation in the activity of searching for a word. Semiotica 62, 51–75.

Gorisch, J., Wells, B., Brown, G., 2012. Pitch contour matching and interactional alignment across turns: An acoustic investigation. Lang. Speech 55, 57–76.

Gravano, A., Hirschberg, J., 2011. Turn-taking cues in task-oriented dialogue. Comput. Speech Lang. 25, 601–634.

Hain, T., Burget, L., Dines, J., Garner, P., Grezl, F., Hannani, A., Huijbregts, M., Karafiat, M., Lincoln, M., Wan, V., 2012. Transcribing meetings with the AMIDA systems. IEEE Trans. Audio Speech Lang. Process. 20, 486–498.

Heldner, M., 2011. Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. J. Acoust. Soc. Amer. 130, 508–513.

Heldner, M., Edlund, J., 2010. Pauses, gaps and overlaps in conversations. J. Phonet. 38, 555–568.

Hjalmarsson, A., 2011. The additive effect of turn-taking cues in human and synthetic voice. Speech Comm. 53, 23–35.

Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., Wooters, C., 2003. The ICSI meetings corpus. In: Proc. ICASSP-2003, Hong Kong, pp. 364–367.

Jefferson, G., 1983. Two explorations of the organisation of overlapping talk in conversation, 1: Notes on some orderliness of overlap onset. Tilburg Pap. Lang. Liter.

Jefferson, G., 1987. Notes on 'latency' in overlap onset. In: Button, G., Drew, P., Heritage, J. (Eds.), Interaction and Language Use: Special Issue of Human Studies, vol. 9, pp. 153–183.

Jefferson, G., 2003. A sketch of some orderly aspects of overlap in natural conversation. In: Lerner, G. (Ed.), Conversation Analysis: Studies from the First Generation. John Benjamins Publishing Company, Amsterdam.

Jefferson, G., 2004. Glossary of transcript symbols with an introduction. In: Lerner, G. (Ed.), Conversation Analysis: Studies from the First Generation. John Benjamins Publishing Company, Philadelphia, pp. 13–23.

Kendon, A., 1967. Some functions of gaze direction in social interaction. Acta Psychol. 26, 22–63.

Kurtić, E., Brown, G.J., Wells, B., 2009. Fundamental frequency height as a resource for the management of overlap in talk-in-interaction. In: Barth-Weingarten, D., Dehe, N., Wichmann, A. (Eds.), Where Prosody Meets Pragmatics, . In: Studies in Pragmatics, vol. 8, pp. 183–205.

Kurtić, E., Brown, G.J., Wells, B., 2010. Resources for turn competition in overlap in multi-party conversations: Speech rate, pausing and duration. Proc. Interspeech 2010. Makuhari, Japan.

Kurtic, E., Wells, B., Brown, G., Kempton, T., Aker, A., 2012. A corpus of spontaneous multi-party conversation in Bosnian Serbo-Croatian and british English. In: Proc. Internat. Conf. on Language Resources and Evaluation LREC 2012, Istanbul, Turkey.

Lee, C., Lee, S., Narayanan, S., 2008. An analysis of multimodal cues of interruption in dyadic spoken interactions. In: Proc. Internat. Conf. on Spoken Language Processing (INTERSPEECH'08), Brisbane, Australia.

Lerner, G., 1999a. Collaborative turn sequenes. In: Lerner, G. (Ed.), Conversation Analysis: Studies from the First Generation. John Benjamins Publishing Company, Philadelphia.

Lerner, G., 1999b. Turn-sharing: The choral co-production of talk of interaction. In: Fox, B., Thompson, S. (Eds.), The language of Turn and Sequence. Oxford University Press, Oxford/New York, pp. 225–256.

Levinson, S., 2006. On the human 'interaction engine'. In: Enfield, N., Levinson, S. (Eds.), Roots of Human Sociality: Culture, Cognition and Interaction. Berg, Oxford/New York, pp. 39–69.

Local, J., 2005. On the interactional and phonetic design of collaborative completions. In: Hardcastle, W., Beck, J. (Eds.), A Figure of Speech: A Festschrift for John Laver. Lawrence Erlbaum Associates Inc., pp. 263–282.

Local, J., Walker, G., 2005. Metholodogical imperatives for investigating the phonetic organisation and phonological structures of spontaneous speech. Phonetica 62, 120–130.

Mondada, L., Oloff, F., 2011. Gestures in overlap: The situated establishment of speakership. In: Stam, G., Ishino, M. (Eds.), Integrating Gestures: The interdisciplinary nature of gesture. John Benjamins, Amsterdam.

Quinlan, R., 1994. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.

Reidsma, D., de Kok, I., Neiberg, D., Pammi, S., van Straalen, B., Truong, K., van Welbergen, H., 2011. Continuous interaction with a virtual human. J. Multimodal User Interf. 4, 97–118. http://dx.doi.org/10.1007/s12193-011-0060-x.

Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A simplest systematics for the organisation of turn-taking for conversation. Language 50, 696–735.

Schegloff, E., 1982. Discourse as an interactional achievement: Some uses of uh huh and other things that come between sentences. In: Tannen, D. (Ed.), Georgetown University Round Table on Languages and Linguistics (GURT 1981), . In: Analysing Discourse: Text and Talk. Georgetown University Press, Washington, DC, pp. 71–93.

Schegloff, E., 1987. Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In: Button, G., Lee, J.R.E. (Eds.), Talk and Social Organisation. Mulitlingual Matters, Clavendon, UK, pp. 70–85.

Schegloff, E., 2000. Overlapping talk and the organisation of turn-taking for conversation. Lang. Soc. 29, 1–63.

Schegloff, E.A., 2001. Accounts of conduct in interaction: Interruption, overlap and turn-taking. In: Kaplan, H.B., Turner, J.H. (Eds.), Handbook of Sociological Theory, . In: Handbooks of Sociological and Social Research. Springer, pp. 287–321.

Selting, M., 1998. TCUs and TRPs: The construction of units in conversational talk. Interaction and Linguistic Structures 4, 1–48.

Shriberg, E., Stolcke, A., Baron, D., 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, New Jersey.

Shriberg, E., Stolcke, A., Baron, D., 2001b. Observations on overlap: Findings and implications for automatic processing of multi-party conversation. Proc. 7th European Conf. on Speech Communication and Technology (EUROSPEECH '01). Aalborg, Denmark.

Sidnell, J., 2001. Conversational turn-taking in a Caribbean English creole. J. Pragmatics 33, 1263–1290.

Stivers, T., 2008. Stance, alignment and affiliation during storytelling: When nodding is a token of affiliation. Res. Lang. Soc. Interac. 41, 31–57.

Szczepek-Reed, B., 2006. Prosodic Orientation in English Conversation. Palgrave Macmillan, Basingstoke, UK.

Walker, M.B., Trimboli, C., 1982. Smooth transitions in conversational interactions. J. Soc. Psychol. 117, 305–306.

Wells, B., Corrin, J., 2004. Prosodic resources, turn-taking and overlap in children's talk-in-interaction. In: Couper-Kuhlen, E., Ford, C.E. (Eds.), Sound Patterns in Interaction. John Benjamins, Amsterdam, pp. 119–144.

Wells, B., McFarlane, S., 1998. Prosody as an interactional resource: Turn-projection and overlap. Lang. Speech 41, 265–294.

Yngve, V., 1970. On getting a word in edgewise. In: Papers from the 6th Regional Meeting of the Chicago Linguistic Society, pp. 567–577.

Please cite this article in press as: Kurtić, E. et al., Resources for turn competition in overlapping talk, Speech Comm. (2013), http://dx.doi.org/10.1016/j.specom.2012.10.002