

F0 accommodation and turn competition in overlapping talk

Emina Kurtić¹, Jan Gorisch*

Institute for the German Language (Institut für Deutsche Sprache), Mannheim, Germany

Abstract

To date, little is known about prosodic accommodation and its conversational functions in instances of overlapping talk in conversation. A major conversational action that happens in overlap is turn competition. It is not known whether participants accommodate prosodic parameters locally in the overlapped turn (initialisation) or access a repertoire of prosodic patterns that refer to general prosodic parameter norms (normalisation) when competing for the turn in overlap. This paper investigates the initialisation and normalisation of fundamental frequency (f0) and assesses its role as a resource for turn competition in overlap. We drew instances of overlapping talk from a corpus of conversational multi-party interactions in British English. We annotated the overlaps on a competitiveness scale and categorised them by overlap onset position and conversational function. We automatically extracted f0 parameters from the speech signal and processed them into f0 accommodation features that represent the normalising or the initialising use of f0. Using decision tree classification we found that f0 accommodation is only relevant as a turn competitive resource in overlaps that start clearly before a speaker transition. In this turn context, we found that normalising and initialising f0 features can both be relevant turn competitive resources. Their deployment depends on the conversational function of overlap.

Keywords: F0 accommodation, Interactional phonetics, Decision tree modelling, Talk-in-interaction, Overlapping talk, Turn competition, Pitch contour matching

1. Introduction

There is now substantial research evidence to support the observation that participants in conversation accommodate the prosodic design of their turns in talk to each other. Prosodic accommodation is found to be related to increasing conversational harmony (Lakin & Chartrand, 2003), decreasing misunderstanding and ensuring conversational success (Pickering & Garrod, 2004; Parrill & Kimbara, 2006; Pickering & Garrod, 2006), displaying a positive attitude towards the conversational partners and achieving acceptance with a social group (Chartrand & Bargh, 1999) or acknowledging social hierarchies within a group (Giles, Coupland, & Coupland, 1991). Furthermore, an increase in conversational involvement (De Looze, Oertel, Rauzy, & Campbell, 2011) and higher ratings of conversation friendliness (Zellers & Schweitzer, 2017) have been associated with prosodic accommodation between participants.

Overlapping talk is a pervasive phenomenon in conversations (Çetin & Shriberg, 2006; Kurtić, Brown, & Wells, 2013; Sacks, Schegloff, & Jefferson, 1974), and it is the most common environment in which speakers compete for the speaking turn (Kurtić, 2011). As speakers negotiate their way through overlap it could be expected that some prosodic accommodation between speakers would take place. Speakers could also accommodate the prosody to their own talk just preceding overlap to mark turn competition. However, to date little is known about prosodic accommodation in overlap. So far overlap has purposefully been left out of studies on prosodic accommodation (De Looze et al., 2011; Heldner, Edlund, & Hirschberg, 2010; Zellers & Schweitzer, 2017). Likewise, very few studies on overlap explicitly consider prosodic accommodation. There are some indications in Conversation Analysis (CA) literature on overlap that prosodic accommodation between speakers could be found in overlap and might be a resource for overlap resolution (Schegloff, 2000) and turn competition (French & Local, 1983) in single overlap instances. Nevertheless, so far most quantitative research on prosody of overlap in conversation (e.g. Lee et al., 2010; Kurtić et al., 2013; Kurtić, Brown, & Wells, 2009) has neglected the possibility of prosodic accommodation in the local overlap context or its between-speaker aspect.

* Corresponding author at: Institute for the German Language, R 5, 6-13, 68161 Mannheim, Germany. Fax: +49 621/1581 200.

E-mail addresses: e.a.kurtic@gmail.com (K. Emina), gorisch@ids-mannheim.de (G. Jan).

¹ Currently without affiliation.

In this paper we aim to fill in this gap and investigate prosodic accommodation as a resource for turn competition in overlap in naturally occurring conversations. In focusing on prosodic accommodation as a resource for turn competition in overlap we contribute to the body of quantitative studies on overlap and turn competition (e.g. Lee et al. (2010), Kurtić et al. (2013)) and elaborate on the primary qualitative findings reported by French and Local (1983) discussed in Section 2. We do not address the question of temporal development of overlaps or the role of prosody in overlap resolution (Schegloff, 2000), nor do we discuss the relatedness of prosodic design of overlap to the outcome of turn competition, i.e. who wins the turn (Jefferson, 1983a; Schegloff, 2000).

The multitude of studies on prosodic accommodation in conversations has produced a terminological variety when referring to the phenomenon including the terms pitch matching (Gorisch, Wells, & Brown, 2012; Zellers & Schweitzer, 2017), entrainment (Brennan, 1996; Levitan, Gravano, & Hirschberg, 2011) and prosodic orientations (Szczepek Reed, 2006, 2012), which are used in studies most related to this work. In this paper we follow De Looze, Scherer, Vaughan, and Campbell (2014) in the definition of the term. Henceforth, the term *prosodic accommodation* is used as a general term, whereas speakers' accommodating behaviours include *prosodic adaptation*, where speakers' prosodic design of turns becomes more similar, and *prosodic differentiation*, where the turn designs diverge.

Previous studies have considered between-speaker accommodation in terms of a variety of prosodic parameters: vocal intensity (Natale, 1975), fundamental frequency (f0) (or its perceptual correlate pitch) (Couper-Kuhlen, 1996; Gorisch et al., 2012; Szczepek Reed, 2006; Wells & Stackhouse, 2015; Zellers & Schweitzer, 2017), speech rate and speech timing (Lerner, 2002; Matarazzo, Weitman, Saslow, & Wiens, 1963) or their combinations (Müller, 1996; Kousidis et al., 2008; Kousidis, Dorran, McDonnell, & Coyle, 2009; Selting, 2010; De Looze et al., 2011; De Looze et al., 2014). However, it was reported that accommodation in f0 is more frequent in conversations than accommodation in other prosodic parameters (Bonin et al., 2013; De Looze et al., 2014). Furthermore, f0 was found to be a relevant turn competitive resource (Kurtić et al., 2009), as well as to provide contextualisation information that makes turns in conversation interpretable on various levels of meaning (Ogden, 2006). For example Zellers and Ogden (2014) show that f0 matching was found in turn transitions where something problematic is being done, which signals participants' orientation to progressing the conversation and avoiding a break down despite the dissociative action. According to Schegloff (2000) turn competition is such a dissociative action in which participants demonstrate "problematicness" and a need for resolution. Therefore, in this paper we focus on f0 accommodation and its use as a turn competitive resource.

Central to this work are the two views on f0 accommodation introduced by Zellers and Schweitzer (2017): *initialisation* and *normalisation*. Initialising approaches (e.g. Gorisch et al., 2012) analyse f0 in local context of each single turn in talk and claim that speakers accommodate to the co-participants' or their own f0 values in this local context. The normalising approach on the other hand (e.g. Sicoli, Stivers, Enfield, & Levinson, 2015) assumes that speakers are able to estimate

their co-participants' f0 norms and will match those general norms without reference to the local turn context. For the investigation of f0 accommodation in overlap we define as *initialising f0 features* those that describe f0 accommodation to co-participants' or speakers' own values in turns containing and preceding overlap. We contrast them with the *normalising f0 features* that capture speakers' f0 accommodation relative to their individual norms generalised over the entire conversation.

Turn competition almost always takes place in overlap, but not all overlaps are turn competitive. Contrasting the f0 design of competitive and non-competitive overlaps allows us to draw conclusions about the relevance of f0 accommodation as a resource for turn competition. To this end the aims of this study are as follows.

Since to date there is no evidence on the f0 initialisation in relation to turn competition in overlap we first aim to investigate this relationship and ask whether

1. f0 initialisation in turns containing overlap is a relevant resource for turn competition.

If this is so, it will be possible to discriminate between turn competitive and non-competitive overlaps based on initialising f0 features more accurately than it would be possible when no f0 information at all is used.

Next, we aim to evaluate the relevance of f0 initialisation in turn competition relative to f0 normalisation, which is known from previous work (Kurtić et al., 2009) to take place in turn competition. It is possible that either or both of the f0 accommodation configurations is used by the participants when competing for the turn. This leads to the following possible outcomes:

2. f0 normalisation rather than f0 initialisation is a turn competitive resource in overlap
3. f0 initialisation rather than f0 normalisation is a turn competitive resource in overlap
4. f0 accommodation in terms of both f0 normalisation and f0 initialisation takes place when speakers compete for the turn in overlap.

In case of 2 we expect that f0 normalisation features will significantly better differentiate between turn competitive and non-competitive overlaps than f0 initialisation features. If 3 holds the opposite will be the case. Finally, if 4 holds, we expect that initialising and normalising f0 features in combination will discriminate between competitive and non-competitive overlaps significantly better than no f0 information at all.

Investigating these possible uses of f0 accommodation in turn competition has an implication for the relationship between prosodic constructions and their interactional meanings.² The relevant question here is: Are prosodic constructions mapped to their interactional meanings before being employed in conversation, or do they emerge in a case-by-case fashion as conversation participants co-create them in the course of a conversation? An evidence for the normalising view on prosodic accommodation (2) would suggest that there is a mapping

² We use the term *interactional meaning* in this work to refer to the interpretation shared by the participants of what a prosodic construction is doing in conversation (Kelly & Local, 1989). This level of meaning is also referred to as discursive meaning (Cruttenden, 1997), speech act (Levinson, 1983), social action (Couper-Kuhlen & Selting, 2006), dialogue act (Core & Allen, 1997; Shriberg, Dhillon, Bhagat, Ang, & Carvey, 2004; Bunt, 2011) or simply conversational function (De Looze et al., 2014; Sicoli et al., 2015; Bunt, 2011). We adopt the proposal in Local (2003) on how these meanings are realised by prosodic domains and features.

between prosodic patterns and their interactional meanings which happens outside the conversational context. An evidence for the initialisation view on prosodic accommodation (1 and 3) on the other hand would suggest that prosodic patterns in conversation are co-constructed by the participants at the point in conversation at which an overlap occurs, so prosodic patterns have no interactional meanings per se, but only at these given points in conversation. A third conclusion is also possible: Prosodic patterns are mapped to recognisable interactional meanings outside the local turn context, but they might need to be modified in individual overlaps by reference to the prosodic design of the turns containing these overlaps. This would be the case if we can retain (4) and is a likely conclusion given that there is evidence in the literature for both the normalising and initialising approach to f0 accommodation, as well as the results in Zellers and Schweitzer (2017) as discussed in Section 2.

To address these questions we adopted a method that intersects between Conversation Analysis (CA), Interactional Phonetics (IP) and Data Mining. We first performed an interactional categorisation task and annotated 3043 overlap instances as turn competitive or non-competitive. Second, we extracted initialising and normalising f0 features as well as further non-prosodic features (type of overlap and the position of an overlap onset within the ongoing turn) that might constrain the use of f0 accommodation features for turn competition. Finally, decision tree classification was used to analyse the relationship between f0 accommodation and turn competition.

In the following we first review the relevant previous research (Section 2) and explain the data and the method we adopted in this study (Section 3). In this section we describe the interactional categorisation of overlaps and its evaluation for inter-annotator agreement, followed by the phonetic analysis, including a discussion of features to be analysed, the feature extraction, their reliability and the classification experiment. Section 4 presents the results of the classification experiments followed by a discussion of the results in Section 5 and conclusion in Section 6.

2. Related work: initialisation, normalisation and the study of overlap

Zellers and Schweitzer (2017) group the approaches to studying f0 accommodation in conversation into two categories: initialisation and normalisation. As noted above, initialisation approaches regard f0 accommodation as a local phenomenon that takes place between adjacent turns in talk. Normalising approaches on the other hand assume that participants have a model of each other's f0 norms and estimate f0 patterns as they occur in turns relative to these general norms without a reference to the local context.

All previous studies on f0 and its meaning in or out of the conversational context could be seen as adopting one or both of these approaches. However, it was not until Zellers and Schweitzer (2017) that both these approaches have been explicitly evaluated in relation to studying f0 accommodation and its interactional meaning. As Zellers and Schweitzer (2017) note, the literature is rich with evidence for both the initialising and the normalising view on f0 accommodation.

Much of the previous work on the meaning of intonational features in conversation that would be compatible with the normalising view, does not consider f0 initialisation at all. This includes studies on the meaning of f0 contours in conversation that lean on the autosegmental-metrical framework. Hedberg, Sosa, Görgülü, and Mameni (2010) for example propose nine categories of interactive function of different tones (in the ToBI framework) found in American English WH-questions: Elaborative Detail, Floor Deferring, Directing Information Flow, Rhetorical, Supplementary Information, Topic Initiator, Reciprocal Question, Clarification and Return to Old Topic. Calhoun and Schweitzer (2012) propose that f0 contours frequently associated with words and short phrases are stored in the mental lexicon along with their discourse meanings suggesting that local f0 accommodation is not taking place when these phrases are used in conversation.

Furthermore, quantitative studies on meaning of prosody in conversation which do not work with f0 categories, but with the speech signal itself, e.g. Shriberg & Stolcke (2004), measure f0 events relative to the same speaker's f0 realisation in the conversation as a whole or in selected contexts.

On the other hand studies on prosodic accommodation in which the degree of accommodation between speakers is investigated in relation to its conversational functions and/or distribution patterns across the conversation could be seen as initialising in nature (De Looze et al., 2011, 2014; Lee et al., 2010; Heldner et al., 2010; Kousidis et al., 2008; Kousidis et al., 2009; Levitan & Hirschberg, 2011). In these studies prosodic features of moving windows of conversation (Kousidis et al., 2009), naturally occurring turns (Lee et al., 2010; Heldner et al., 2010) or a combination of both (De Looze et al., 2014) are computed for each speaker and then related between speakers by correlation or other metrics (Burgoon, Stern, & Dillman, 2007) before determining the conversational function of identified f0 accommodation patterns.

The work by Kousidis et al. (2008, 2009) shows that participants adapt their prosodic features on a relatively long time-scale from 10 s up to 1 min. Additionally, participants tend to match each other prosodically not only on a long but also on a short time-scale, e.g. in the environment of backchannels, as has been demonstrated by Levitan et al. (2011). De Looze et al. (2014) find that f0 accommodation between speakers in terms of synchrony or adaptation as we refer to it here, leads to conversations perceived as more natural with speakers being more engaged and liking each other more. A differentiation in f0 or asynchrony on the other hand leads to a perceived disbalance in speaker engagement with one of the participants being seen as dominating the conversation. Lee et al. (2010) report higher f0 adaptation in interactions between married couples that are rated as positive. In addition, the prosodic accommodation features extracted across their conversation sessions allow to accurately classify the entire sessions as positive or negative.

In some studies on f0 accommodation it is even explicitly stated that an initialising prosodic analysis is preferable to a normalising one. Heldner et al. (2010) for example claim that local prosodic models are more suitable for characterising the way backchannels differ from other vocalisations than general models. The way a backchannel differs from just

preceding talk of a speaker is more robust than estimations computed relative to the overall conversational norms.

The initialising approach to f_0 accommodation is also found in studies that investigate conversational data and either completely or partly adopt the methodological framework of Conversation Analysis (CA) and Interactional Phonetics (IP). This is because these disciplines put the orientations of conversation participants to each other's talk as a central piece of evidence for any analytic claim.

Gorisch et al. (2012) investigate the prosodic design of interactionally aligned and non-aligned insertions, i.e. short turns inserted between another speaker's turns, using a metric that combines f_0 and intensity. They find that insertions in which speakers align with respect to conversational action are significantly more similar than non-aligning ones. Since alignment is accompanied by prosodic similarity they conclude that prosodic accommodation is a phenomenon that takes place in a local turn context.

The CA work on overlapping talk and turn competition also contains some hints about the prosodic accommodation being performed in the local overlap context, compatible with the initialising view on prosodic accommodation. Schegloff (2000) introduces an overlap resolution device as composed of (1) turn-competitive "hitches and perturbations" or cues, (2) the places in or the phases of the overlap at which these cues are deployed and (3) the mechanism that describes how an overlap develops and ends (Schegloff, 2000, p. 11). He identifies higher pitch and loudness as prosodic cues for turn competition in overlap in addition to pace, cut-offs, sound stretches and repetitions of prior material. According to Schegloff, speakers process overlaps beat by beat, with beats roughly corresponding to syllables. If speakers are competing for the turn, they start using the turn-competitive cues at latest on the second beat. For recipients this is also a cue to turn competition, whereupon they can decide whether to enter the competition or to withdraw.

This explanation of beat-by-beat development of overlap points to participants' ability to precisely monitor each others talk in and around overlap and to respond to each other by accommodating prosodic cues over units of a very short duration to mark turn competition. Schegloff is less explicit about the nature of this prosodic accommodation and states that "each (participant) can shift to 'competitive production' by deploying any of the hitches and perturbations ..." (Schegloff, 2000, p. 21). However, his explanation of turn competitive cues suggests that he might advocate an initialising view on prosodic accommodation because he states that "talk can get suddenly (i) louder in volume, (ii) higher in pitch, ..." (Schegloff, 2000, p. 12) indicating that a change happens relative to the preceding talk.

CA and IP studies, however, do not uniquely advocate the initialising approach. Instead some offer evidence that f_0 normalisation is a preferred analysis.

Szczepek Reed (2006) for example reports on prosodic complementation and continuation, in which interlocutors take up and complete or continue a previous intonational contour (and occasionally rhythm). This suggests that both speakers have a clear notion of what the continuation of the intonational contour should be. Likewise, Sicoli et al. (2015) find that a conversational function of a question can be determined by its initial f_0 design only regardless of the preceding turn's prosodic design.

Some insights on phonetic resources for turn competition in overlap are also based on the data that derives from the normalising view on f_0 accommodation. Kurtić et al. (2013) report that f_0 and intensity features in overlap regions normalised to the speakers' overall values for non-overlapped turns are turn competitive cues. Kurtić et al. (2009) present a detailed investigation of f_0 in turn competitive and non-competitive overlaps. They analyse f_0 of each speaker involved in overlap, over-lapper and over-lappee, however, the in-overlap f_0 is compared to each speaker's norms for non-overlapped turn beginnings and do not consider any modification in the local context of overlap. They find that overlappers raise f_0 relative to their norms to initiate a turn competitive overlap, and over-lappees do the same if they return competition.

On the other hand there is also evidence that both normalisation and initialisation are at work for the realisation of turn competition in overlap. French and Local (1983) describe the use of prosodic resources for turn competition as follows: "... the portion is markedly raised in pitch and loudness respectively. By 'raised' we intend the following: 1. that the subscripted portion is both higher and louder than that speaker's norm for beginning turns at points where another speaker has completed his turn 2. that it is both higher and louder than any portion of incomer's speech not so marked 3. that it is loud, but not necessarily high in absolute terms, relative to the speech contained in current turn." (French & Local, 1983, p. 23).³

In this work the modification of phonetic parameters that do turn competition in overlap is not only relative to an over-lapper's norm for out of overlap talk (1 and 2), but also relative to those of the interlocutor (3). However, despite considering prosodic initialisation French and Local (1983) adopt a normalising view on *pitch* accommodation: definition (3) of high and loud above suggests that there is accommodation between speakers in intensity only and that pitch (high) is only realised relative to an incomer's own norms for turns in general (1 and 2).

Finally, some interactional studies suggest that both initialisation and normalisation are relevant *and* carry different interactional meanings. Couper-Kuhlen (1996) for example suggests that the interactional meaning of a turn can determine whether f_0 normalisation or initialisation is used. Her study reports on f_0 accommodation in word repetitions at openings of telephone calls at a radio phone-in program. She finds that participants use similar pitch levels but relative to their own ranges when "quoting", which is an affiliative action. On the other hand exactly same pitch levels irrespective of the speakers' range are used when participants are "mimicking", which can be disaffiliative. Therefore, quoting requires f_0 normalisation to speakers' own range, but mimicking does not. Instead it is an exact copy of the co-participants' f_0 level at a given turn in talk, which would comply with an initialising view.

The findings in Zellers and Schweitzer (2017) also suggest a similar conclusion. The authors aim to evaluate which approach offers more accurate metrics for studying f_0 accommodation and evaluate both initialisation and normalisation on three tasks: (1) participants' ratings of each other's friendliness and likeability, (2) modality of conversation (face-to-face vs. blind) and (3) syntactic form of adjacent turns (question-answer, vs.

³ This same view is also adopted in Wells and Macfarlane (1998), another IP study on turn competition in overlap.

statement-statement). The results for all three tasks differ. While initialising f0 seems to be more relevant for friendliness ratings, normalising f0 gave a better model for the syntactic form distinction and there was no significant difference in the two approaches related to the modality of conversation. The authors conclude that it might be the task at hand that makes one or another approach more suitable for the study of f0 accommodation. Essentially, this indicates that at least until further evidence is available each study on f0 accommodation needs to evaluate both approaches simultaneously in order to know which one is suitable for the task at hand.

The aims of this work as set out above require such a simultaneous investigation of f0 initialisation and normalisation for turn competition in overlap. Therefore this work adds to the discussion of these two approaches to studying f0 accommodation in conversation recently opened by Zellers and Schweitzer (2017). Our study also completes the currently available evidence on the role of normalising f0 in turn competition by adding the initialising view on f0 accommodation. As such it fills the gap in the existing work on prosody of turn competition in overlap and contributes to the broader enterprise of interactional phonetics, which describes how different prosodic patterns are systematically used and oriented to by conversational participants in order to implement social actions. Finally, as described above the investigation of the initialising and normalising views on prosodic accommodation has a further implication for understanding the relationship between prosodic constructions and their interactional meanings, and this study offers further evidence to this end.

3. Material and method

3.1. Data

3.1.1. Corpus

Our analyses are based on overlaps drawn from a corpus of multi-party face-to-face conversations (3 h long) between four friends speaking British English (Kurtić, Wells, Brown, Kempton, & Aker, 2012). These were freely flowing conversations without a task or a pre-assigned topic, which were recorded on video and audio in a quiet room. The audio was recorded via head-mounted microphones for each speaker on a separate audio channel. Each speaker's microphone was pointed away from the neighbouring speaker. This gives a very high quality sound and has the advantage that parameters such as pitch can be extracted reliably per participant even within overlap regions. The audio recordings have been manually segmented into Turn Constructional Units (TCUs) following Sacks et al. (1974). TCUs are conversational units that have points of completion projectable by the conversation participants, called Transition Relevance Places (TRPs).⁴ At a TRP, speaker change can occur but does not have to. Therefore speaking turns can consist of one or more TCUs.

⁴ Transition Relevance Place (TRP) is defined as a possible place of turn completion (Sacks et al., 1974). It is the place in turn where the speaker transition becomes relevant although it is not necessarily accomplished (Schegloff, 1996, p. 55). Jefferson (1983a) speaks about TRP space around a possible completion point rather than understanding a TRP as one point in turn. In this work TRP refers to a TRP space in the sense of Jefferson (1983a) (see also Kurtić (2011, 26–35) for a more detailed explanation of Jefferson's overlap onsets).

3.1.2. Overlap collection

Overlap instances were detected automatically using the start time (ST) and end time (ET) information for each TCU. The multi-party nature of the data allows for multiple speakers to overlap simultaneously. These so called multi-speaker overlaps have been found to be typically very short (Kurtić, 2011) and are excluded from observation in this study. We only focus on overlaps between two speakers and exclude all instances in which the voice of a third party is present. For a speaker pair (A, B) an overlap instance is defined if and only if $ST(TCU_A) \leq ST(TCU_B) \leq ET(TCU_A)$, where $ST(TCU_{speaker})$ is the start time of a speaker's TCU and $ET(TCU_{speaker})$ is the end time of a speaker's TCU. In this case, speaker A's TCU is overlapped and therefore speaker A is called **overlappedee**. Speaker B overlaps speaker A's turn and is therefore called **overlapper**.

This automatic method of overlap detection may include overlaps of very short duration, that are practically not heard as overlaps. This can be caused by faulty speech non-speech segmentation at the transcription stage of the corpus. These cases have been marked during annotation and we exclude them from the analysis. Furthermore, overlaps that result from two parallel conversations, like asides or conversational schisming (cf. Sacks et al., 1974; Schegloff, 2000) are not considered here either. Exclusion of these cases results in 3043 overlap instances that we use for the analysis.

3.2. Interactional analysis

The main aim of the interactional analysis in this study is to assess the competitiveness of each of the 3043 overlap instances in our data set and further classify these overlaps into types identified by previous work. Additionally, our interactional analysis includes the categorisation of overlaps according to the so called overlap onset types. This indicates the positioning of the overlap with respect to the ongoing turn and its conversational function and may constrain the use of prosody for turn competition. These two tasks are best approached by the method of Conversation Analysis (CA) because it allows us to consider each overlap in its immediate sequential context and decide about its competitiveness and onset type based on the observable actions of the participants in that sequence.

According to CA, talk is organised in social actions, such as offers, requests, greetings, agreements, etc. (Ten Have, 2007). These are negotiated between the participants on a turn-by-turn basis. Sequences of actions emerge and often come in pairs with first and second pair-part, such as invitation-acceptance or invitation-rejection. Within the CA framework, the identification of the social action is to be analysed according to the treatment of the interlocutors' turns by each other. For example, the social action of an invitation can be identified by the acceptance of the invitation or its rejection. In turn, the social action of acceptance or rejection can be identified by the analysis of what comes next and so forth. The central question is at each analysis step: how do the participants treat their turns, also referred to as the *next turn proof procedure* (Hutchby & Wooffitt, 2008, p. 45).

Using CA to assess overlap competitiveness distinguishes our work from studies like Chowdhury, Danieli, and Riccardi (2015) or Truong (2013), who classify overlap competitiveness

based on annotators' intuitions about overlap competitiveness. This requires a substantial interpretative effort by the annotators, which is possibly reflected in the relatively low inter-annotator agreement scores. Truong (2013) for example report an agreement of (Krippendorff's) $\alpha = 0.30$, while Kurtić (2011) reports that agreement can increase substantially beyond this level with experience and training in analysis of conversational sequences according to the methods of CA.

In the following we describe the competitiveness and overlap type analysis and the resulting annotation of the data set.

3.2.1. Overlap competitiveness

We adopt Schegloff's (Schegloff, 2000) notion of competitiveness in which either speaker can demonstrate the aim to win the current turn. For this reason, the instruction to our annotators was based on the following definition of turn competition:

Definition 3.1. An instance of overlapping speech is competitive if either party, overlappee or overlapper, or both demonstrates the aim to prevent the other party from either keeping or taking over the current turn.

The data set used in this study was annotated by two annotators with the same level of experience in CA. Each annotator analysed the entire set of 3043 overlaps based on guidelines from previous work on overlapping talk (Kurtić, 2011).

Binary vs. graded annotation of competitiveness. The annotators were instructed to indicate the level of competitiveness of each overlap instance on a 5-point scale that describes the competitiveness range as shown in Table 1. In most previous work (Adda-Decker et al., 2008; Chowdhury et al., 2015; Kurtić, 2011; Kurtić et al., 2013; Lee et al., 2010; Oertel, Włodarczyk, Tarasov, Campbell, & Wagner, 2012; Truong, 2013) turn competition has been regarded as a binary attribute, i.e. participants would either compete for the turn or not. For this annotation, we decided to leave the option open that competitiveness may be a more nuanced phenomenon, where clearly competitive and non-competitive cases can be differentiated from those that are less clear. The scale is used to push the annotators to differentiate between clear and not so clear cases of overlap competitiveness. If results arrange around 1 and 5, it would be a good evidence of competitiveness being a binary notion. However, if a systematic spread can be observed, then we would have material for a finer grained investigation into competitiveness nuances.

Table 2 shows the distribution of competitiveness annotations on the 5 point scale by both annotators. The frequencies in Table 2 indicate that "clear" categories are more frequently chosen by the annotators than the "rather" categories.

Table 1
Competitiveness annotation scores ordered on a scale from 1 to 5.

Score	Description
1	Clearly non-competitive
2	Rather non-competitive
3	Neither competitive, nor non-competitive
4	Rather competitive
5	Clearly competitive

Table 2

Distribution of the competitiveness categorisations by annotators A and B according to the competitiveness scale from 1 to 5. There are missing instances in both annotators' data sets. These are discarded here, so the total number of overlaps of 3043 is reduced to 3025 for annotator A and 2838 for annotator B. The percentages in brackets are given relative to these new totals.

Rating	Description	#inst A	#inst B
(1)	clearly non-competitive	1995 (66%)	1966 (69.3%)
(2)	rather non-competitive	216 (7.1%)	202 (7.1%)
(3)	neither nor	43 (1.4%)	34 (1.2%)
(4)	rather competitive	241 (8%)	160 (5.6%)
(5)	clearly competitive	530 (17.5%)	476 (16.8%)

Further evidence can be found in the agreement of the two annotators on the competitiveness categories and their levels. From the confusion matrix in Table 3 we can see on how many instances of each competitiveness category the annotators agree and disagree. If competitiveness is graded, then the "rather" category should be very obvious to the annotators. This would be reflected in their agreement: The agreement on the "rather" category should be as high as that on the "clear" category. As Table 3 shows, however, there are 1932 (82.7%⁵) cases of overlap, for which both annotators agree that they are "clearly" competitive (both annotators rated 5) or not (both annotators rated 1) and only 76 (3.23%) cases in which annotators agree that an overlap is "rather" competitive (both annotators rated 4) or not (both annotators rated 2). There are also 346 (14.7%) cases in which annotators disagree on the "clearness" of competitiveness (one annotator rates 1 or 4 and the other one 2 or 5 respectively). The "rather" category therefore seems much less obvious to the annotators than the "clear", indicating that turn competitiveness might be better understood in terms of a binary distinction than the graded one.

Given these results we decided to follow the previous work and treat overlap competitiveness as a binary phenomenon. For this purpose we collapse the ratings 1 and 2 into a single rating "non-competitive" and 4 and 5 into "competitive". As to the rating 3, we decided to treat it as if it was agreed on as competitive or not if at least one annotator has marked it as such. We discarded the 2 instances for which it is agreed that they belong to category 3.

Inter-annotator agreement. We measured the inter-annotator agreement between the annotators using Krippendorff's α (Krippendorff, 2004). The inter-annotator agreement gives an indication of how well the instructions of the researchers map onto the data in general, i.e. how well the criteria for the categorisation match with observations in the real world. Also, the cases, where the annotators agree and disagree on the competitiveness category provide us with the set of clear and unclear cases respectively. Clear cases are more likely to show a relationship between prosodic features and competitiveness categories than the entire collection of instances. A few examples with accompanying detailed analyses can be found in Appendix A.

The inter-annotator agreement for the data set as originally annotated on the ordinal competitiveness scale of 1–5 is $\alpha = 0.64$. However, as described above the data set has been

⁵ Percentages are calculated relative to the total number of overlaps in which annotators agree on competitiveness, but not on how clear competitiveness is. These are found in the upper left square and the lower right square in Table 3. This count of overlaps is 2354.

Table 3

Confusion Matrix for annotator A and annotator B. The counts consider missing instances in both annotators' data sets (in total 223 instances).

		Annotator B					Total
		1	2	3	4	5	
Annotator A	1	1636	80	15	41	51	1823
	2	131	29	2	28	20	210
	3	18	6	2	6	10	42
	4	55	32	7	47	97	238
	5	111	54	8	38	296	507
Total		1951	201	34	160	474	2820

transformed to contain binary competitiveness classes. This new data set with binary competitiveness is shown in Table 4. There are 2818 instances in total, of which 2426 (86.09%) are agreed on as competitive or not (1917 non-competitive and 509 competitive) and 392 (13.91%) are disagreed upon. The agreement score on this modified data set is $\alpha = 0.63$.

For the analyses reported in the remainder of this paper, we decided to work only on the 2426 agreed cases.

3.2.2. *Overlap onset types*

Overlap onset types is a feature that characterises overlaps in terms of the positioning of the overlap onset relative to the ongoing turn and also indicates why an overlap occurs. The feature values include Jefferson's overlap onset categories, *blind spot*, *terminal*, *progressional* and *recognitional* (Jefferson, 1983b, 1986) and two further values, (*turn incursive* and *simultaneous start*) that describe the positioning of the overlap onset relative to the completion of the turn. Table 5 shows the definitions of overlap onset types. For the purposes of our analysis, the annotation of overlap onset types is important as speakers may use different prosodic resources to compete for the turn in overlaps of different onset types (cf. Section 4).

The overlap onset type annotation was performed independently from the competitiveness annotation in a second annotation run upon completion of competitiveness annotation. This means that the annotation decision on overlap onset types did not influence the decision on overlap competitiveness. Overlap onset types have been annotated by only one of the annotators, so we do not report a detailed inter-annotator agreement analysis for these categories as we did above for the overlap competitiveness analysis. The decision not to double annotate overlap onset types is based on the observation reported in previous work that this type of content is easier to annotate and more likely to render high annotator consistency. Kurtić (2011) reports intra-annotator consistency on the overlap onset types as following: terminal overlaps: $\alpha = 0.90$, simultaneous starts: $\alpha = 0.88$, blind spot overlaps: $\alpha = 0.83$, progressional: $\alpha = 0.87$, recognitional: $\alpha = 0.87$.⁶ These intra-annotator agreement values are substantially higher compared to those for overlap competitiveness, which is $\alpha = 0.70$ (Kurtić, 2011, p. 121).

Table 6 shows the distribution of overlap onset types according to their competitiveness in the finally analysed data set (cf. Section 3.3.3).

⁶ The annotation of overlap onset in Kurtić (2011) does not include turn incursive overlaps.

Table 4

Confusion Matrix for annotators A and B when their competitiveness rating scores are grouped into two categories: competitive (C) and non-competitive (NC).

		B		Total
		NC	C	
A	NC	1917	140	2057
	C	252	509	761
Total		2168	649	2818

Table 5

Overlap onset types based on Jefferson (1983c) ordered according to the position relative to the TRP (cf. Kurtić (2011) for detailed description and examples).

Overlap onset type	Description
Terminal	Overlaps in which overlap onset is placed at the last phonological word (or its parts) of the current speaker's turn, where the end of the turn is reliably projected but the TRP is not yet reached (the beginning of the TRP space)
Simultaneous start	Overlaps in which both speakers start simultaneously exactly at the Transition Relevance Place (cf. Section 1)
Blind spot	Overlaps in which there is latency in overlap start: overlappers demonstrate a change from listenership to speakership before the turn holder changes from the state of speakership to listenership; typically placed immediately after the beginnings of turns (the end of TRP space) and not at ends of syntactically complete units
Progressional	Mid-turn overlaps in which a speaker starts upon a disfluency in the current speaker's turn
Recognitional	Mid-turn overlaps that start when a point of semantic adequacy is achieved, i.e. when the overlapper has had enough information as to what the turn is about
Turn incursive	Mid-turn overlaps, placed clearly outside the TRP space, which cannot be categorised as either progressional or recognitional

Most overlaps (840 out of 1839) occur in "terminal" position of the preceding talk, accounting for almost half the data (45.7%). Almost all of them (801 out of 840) are non-competitive. 317 and 386 overlaps occur at "blind spots" or "turn incursively", while their distribution between competitive and non-competitive overlaps seems to be opposed with overlaps at blind spots being non-competitive and overlaps at turn incursive position being competitive. The turn-incursive position seems to be the most prominent position for competitive overlaps with 256 out of 422. 10% of the overlaps start simultaneously after a gap, with a preference for non-competitive overlaps (158 vs. 27). Some (4.3%) of overlaps occur at a progressional position and a few (1.6%) occur at a recognitional position.

3.3. *Instrumental analysis: f0 features*

To test our hypotheses (cf. Section 1) we extract f0 features incorporating the normalising and the initialising approach to f0

Table 6

Overlap onset types distribution according to their competitiveness in the analysed data set (cf. Section 3.3.3).

Overlap onset type	Competitive	Non-competitive	Frequency
Terminal	39	801	840 (45.7%)
Simultaneous start	27	158	185 (10.1%)
Blind spot	65	252	317 (17.2%)
Progressional	26	53	79 (4.3%)
Recognitional	9	21	30 (1.6%)
Turn incursive	256	130	386 (21.0%)
Missing	–	2	2 (0.1%)
Total	422	1417	1839 (100.0%)

accommodation. The basic unit of analysis for all extracted features is the interactionally motivated turn constructional unit (TCU) (Sacks et al., 1974). The TCU boundaries have been segmented during corpus transcription (cf. Section 3.1.1). Normalising features express speakers' f0 relative to their central tendency values found in each recorded session. Initialising features capture f0 relationships in the local overlap context, i.e. within the in-overlap TCU and in the TCU preceding the overlap. Due to the large number of overlap instances and features, all data is extracted automatically from the acoustic signal and parameters thereof.

3.3.1. Extraction and normalisation of basic contours

We extracted the raw fundamental frequency using the YIN pitch detection algorithm (De Cheveigné & Kawahara, 2002). The YIN algorithm also provides a contour for the aperiodicity in the signal, as the solid line in Fig. 1 illustrates schematically. As both contours f0 and aperiodicity are continuous, we used a threshold on the aperiodicity contour to determine which parts of the continuous f0-contour were to be retained for the subsequent extraction of f0 features.

In order to reflect the auditory properties of pitch perception, we scaled the raw f0 logarithmically. For each speaker at each recording session we measured the distribution of pitch values from which we extracted the speaker's mid range (determined by the median) and the speaker's pitch span (determined by the standard deviation). The retained "pitch" contours, which the algorithm used subsequently, were normalised by subtracting the median and dividing by the standard deviation.

3.3.2. Extraction of features

We extracted normalising and initialising f0 features from the raw f0 and aperiodicity contours described above. Additionally, an overlap duration feature is extracted from the timing information of the transcripts. Altogether, there are 45 features.

Normalising features. To obtain the normalising features we first extracted 6 features related to speakers' general individual pitch contours: *slope*, *minimum*, *maximum*, *standard deviation*, *span* and *height*.

In selection of these features we were guided by previous work. Our aim was to get as close as possible to measures of what is described in the literature concerning impressionistic descriptions of pitch properties (cf. Ogden (2006), who took measures by hand from the contours), but in a fully automatic way. Therefore we included measures of central tendency

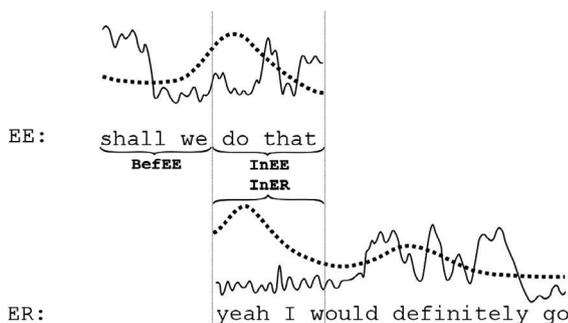


Fig. 1. Schematic illustration of a transcript of two overlapping speakers overlappee (EE) and overlapper (ER) and the basic contours of f0 (dashed line) and aperiodicity (thin solid line). Three contexts of the features are identified through InEE (overlappee in overlap), InER (overlapper in overlap), and BefEE (overlappee before overlap).

such as minimum, maximum, span and height as well as the measures that capture changes in the f0 contour such as overall slope. We could have envisaged more specific f0 features such as "slope on the accented syllable" or "pitch span between the lowest and the highest tone" but this would have required manual labelling of the pitch contour as well, which was beyond the scope of this work.

The resulting normalising features are defined as these 6 features in three overlap contexts: in overlap for the overlapper's talk (*InER*), in overlap for the overlappee's talk (*InEE*) and just before overlap for the overlappee's talk (*BefEE*). As mentioned in the previous section each speaker's feature values are normalised to that speaker's median and SD in each of the recorded sessions.

We also include the *duration* feature, which considers the length of the in-overlap region and the length of the TCU before the overlap. Furthermore, we extract two features: *aperiodAV* and *NanRatio* that describe the periodicity of the signal in the three overlap contexts and allow us to estimate the reliability of the extracted f0 values as described in Section 3.3.3 below. *aperiodAV* is the average of the aperiodicity contour across the overlap regions and *NanRatio* expresses the ratio of missing f0 values to the valid ones. Table 7 summarises the calculation of the feature values.

Initialising features. The initialising features capture the accommodation in f0 in the local overlap context. To this end the pitch and aperiodicity contours are compared in three ways (cf. abbreviations in Fig. 1):

- InIn** Between overlapper and overlappee in overlap.
- InErBef** Between overlapper and overlappee, where the contour of the overlapper is taken from the overlap portion (InER) and the contour of the overlappee is taken from the TCU immediately before overlap (Bef).
- InEeBef** Overlappee (within speaker) contour taken from in-overlap talk and before-overlap TCU.

In order to derive these features, we took the difference of the values from the individual features. For example, $slope_{InErBef} = slope_{ERin} - slope_{EEbef}$ is the comparison of the slopes of the pitch contour of the overlapper in overlap and the pitch contour of the overlappee before the overlap.

These difference based features capture f0 accommodation in terms of *f0 differentiation* (cf. Section 1), i.e. the dissimilarity in the feature values. Another way in which speakers can accommodate their f0 in overlap is *f0 adaptation*, i.e. the turn competition might be realised by speakers f0 becoming more similar.

Table 7
Calculation of feature values.

Feature	Calculation
Slope	Steepness per second (steepness = steepness of the linear polynomial fit to the f0 curve)
f0min	Minimum value of all f0 values
f0max	Maximum value of all f0 values
f0stddev	Standard deviation of all f0 values
Span	f0max-f0min
Height	Median of the contour
Duration	Endtime of Signal (speaker) – starttime of Signal (speaker)
AperiodAV	Mean of aperiodicity contour
NanRatio	Ratio between the number of missing f0 values (due to voicelessness) and the number of valid f0 values

It is not clear whether speakers differentiate or adapt in f_0 when competing for the turn. As outlined above, prosodic adaptation is found to be associated with conversational harmony, friendliness and social concordance. Turn competition is an action in which speakers do the exact opposite, they dissociate from each other as they do it. Indeed this has led Zellers and Schweitzer (2017) to conclude that pitch accommodation in terms of adaptation is not expected during overlapping talk. Therefore it could be dissimilarity or no f_0 accommodation at all that takes place in turn competition. On the other hand f_0 adaptation was shown to be present in turns that implement actions that can potentially be treated as problematic by the participants (Zellers & Ogden, 2014). Turn competition can be seen as such problematic action (Schegloff, 2000), so f_0 adaptation could be the resource to realise it. Therefore in addition to the above measures of difference between speakers, we also capture the similarity between two pitch contours by computing an additional feature: pitch contour similarity or *SimScore*.

Pitch contour similarity (SimScore). We apply an algorithm described by Gorisch et al. (2012), based on dynamic programming (DP). It allows the best alignment path to be found between two time series (here: pitch contours). A by-product of the calculation of the alignment path is the so-called alignment cost, an evaluation of how close the two time series are one to another.

First, the DP algorithm requires a distance matrix (here we used a similarity matrix) where each value (x) of one contour (from speaker A) is compared with each value (y) of the other contour (from speaker B). Formally, we used the following equation following Cooke (1993):

$$\text{sim}(x, y) = e^{-\frac{(x-y)^2}{2\sigma^2}}$$

This allows differences between x and y to be converted into similarities using a Gaussian function (cf. Fig. 2) with a standard-deviation of σ in order to get low scores (close to 0) for dissimilar f_0 -values and high scores (close to 1) for similar f_0 -values. Here we chose $\sigma = 0.3$.

Second, the DP algorithm computes the best alignment path through the distance/similarity matrix, thereby applying dynamic time warping, i.e. it allows different regions of the contours to be aligned in order to minimise the alignment cost. Here, we inverted the procedure of calculating “cost” by calculating “quality” following Aimetti (2011). We further normalised this quality score for missing data (unvoiced regions) and call it here *similarity score (SimScore)*.

This metric does not measure specifically the overall shape of the f_0 contour, the f_0 level, direction of change, rate of change, etc. It rather measures all of these at once.

As with other features the f_0 contours that we compare with *SimScore* come from three contexts: the overlapper (within overlap region: InER), the overlappee (last TCU before the overlap region: BefEE) and the overlappee (also within overlap region: InEE). The *SimScore* features are defined as follows:

SimScore The similarity score between the f_0 contour of the overlapper within the overlap region and the f_0 contour of the overlappee within the overlap region.

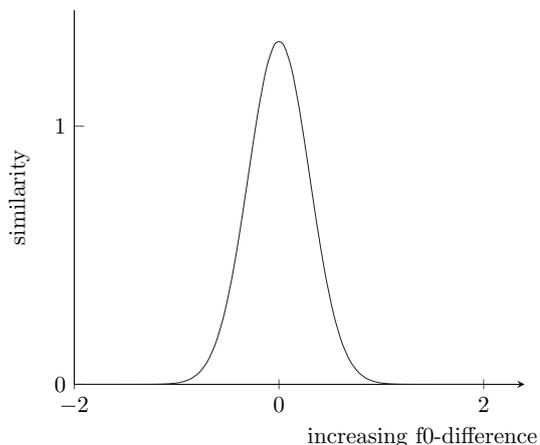


Fig. 2. Gaussian function illustrating the re-distribution of f_0 -differences to similarity.

SimScoreSelfEE The similarity score between the f_0 contour of the overlappee within the overlap region and the f_0 contour of the overlappee (same speaker) before the overlap region.

SimScoreEEbefERin The similarity score between the f_0 contour of the overlappee before the overlap region and the f_0 contour of the overlapper within the overlap region.

3.3.3. The reliability of f_0 features

An analysis based on a set of features extracted in a fully automatic procedure crucially depends on the reliability of the extracted f_0 values. As an indicator of how reliable our f_0 features are we computed the ratio between periodic and aperiodic stretches of the signal (*NanRatio*) for each context. A high *NanRatio* indicates that the proportion of periodicity values is very low in the signal, i.e. if there are no or very few valid f_0 values. As an inspection of a set of examples indicated, the features extracted from these sparse f_0 readings are non-existent or inappropriate for drawing any conclusions in the analysis. Furthermore, in order to compute the co-ordinated features, we need valid feature values in at least 2 out of 3 overlap contexts (BefEE, InEE and InER). If this is not given, individual features would be preferred in the classification, just by the virtue of having more non-missing values.

Therefore, we further clean the data set to include only overlap instances with reliable f_0 estimates. We first remove all overlaps in which no valid f_0 values could be extracted from any 2 or more contexts. Furthermore, we remove those instances with very high *NanRatio* values. An investigation of the *NanRatio* distribution indicates that the removal of all cases with *NanRatio* values higher than the 95-percentile is appropriate as it removes outliers in f_0 features that were due to unreliable f_0 estimates. Furthermore, if the duration of a context is very short (less than 0.3 s for in-overlap (InEE, InER) and less than 0.1s in pre-overlap (BefEE)), there are cases with extreme f_0 values that result from erroneous analysis. We remove all overlaps where this is the case. These measures further reduce our data set for analysis from 2426 overlaps to 1839 overlaps, 1417 (77.1%) of which are non-competitive and 422 (22.9%) are competitive.

3.4. Classification experiment

To evaluate the relevance of f0 accommodation as a turn competitive resource we compare the prosodic design of turn competitive and non-competitive overlaps in terms of f0 features described in the previous section. Differences in the design between these overlap groups allow us to investigate what f0 features make overlaps turn competitive and to test whether f0 accommodation plays a role in turn competition. Contrasting competitive and non-competitive overlaps is best posed as a classification task as previous studies have done (Chowdhury et al., 2015; Kurtić et al., 2013). Given the normalising and initialising f0 features, the task is to classify each overlap instance as turn competitive or non-competitive.

3.4.1. Data: agreed cases

For the overlap classification experiment we use only those cases for which we have both annotators' agreement on overlap competitiveness after collation of competitiveness levels into the binary category-set with the possible values *competitive* and *non-competitive* and after removal of overlap instances for which we cannot assume a reliable feature extraction (cf. Section 3.3.3). This set contains 1839 overlaps, 1417 (77.1%) of which are non-competitive and 422 (22.9%) are competitive. Due to the maximal inter-annotator agreement on this set, we assume it as a gold standard for studying the prosodic design of overlap competitiveness. The disagreed cases are discarded in this analysis.

3.4.2. Experimental set up

Given a training set of human annotated overlaps, the classification algorithm is expected to reproduce the competitive/non-competitive classifications on a previously unseen data set. We perform classifications of overlap competitiveness with the Decision Tree (DT) classifier (C4.5 (Quinlan, 1993)) because of its good performance on our data set, while also offering the possibility to easily interpret the output models, in particular to understand the interactions between features.

We evaluated eight different classifiers from the Weka data mining toolkit (Witten, Frank, Hall, & Pal, 2015) on the overlap competitiveness classification task using f0 features. The results are as follows (*kappa* scores in brackets): Decision Tree C4.5 (0.25), Naive Bayes (0.07^{*†}), Logistic Regression (0.23), SVM (0.26), IBk instance based classifier (0.09^{*†}), LWL instance based classifier (0.19^{*†}), Bagging (0.27), AdaBoostM1 (0.21), Random Forest (0.26). † indicates a performance not significantly different from the majority baseline (0.00). * indicates a significantly lower performance than the best classifier (Bagging). In all other cases there is no significant difference to Bagging. The ensemble learning methods Bagging, AdaBoostM1 and Random Forest all have a Decision Tree as a base learner. These methods do not create easily interpretable models, but are included as they are commonly used to produce more stable models than single decision trees (Zhou, 2012). A similar performance of the single DT compared to its use in the ensembles is an indication that the DT model used here is stable for different data sets.

To maximally exploit the given limited data set, we use the 10-fold stratified cross-validation training/testing strategy. To

evaluate the performance of the classification algorithm we measure the agreement between the automatic and human classifications using the Cohen's kappa metric (Cohen, 1960). The kappa metric is commonly used for measuring inter-annotator agreement and is also recommended for evaluation of automatic classifiers (Japkowicz & Shah, 2011), where chance corrected evaluation of classifiers on imbalanced data sets is required. For our data set, we fix the human-human agreement on competitiveness classification at *kappa* = 1, since only cases in which both human annotators agree are included into the analysis of clear cases (cf. Section 3.4.1). We compute the statistical significance of the differences between classifier and annotators using the two-tailed t-test and report it at the significance level of $p < 0.05$ in all experiments. As baseline we use the majority classifier that classifies all cases as the majority class, in our case the non-competitive overlaps, without using any f0 information and achieves *kappa* = 0.

4. Results

4.1. Competitiveness classification: TRP positions

Previous work (Kurtić et al., 2013) has noted that the prosodic features are only employed as a turn competitive resource if an overlap is initiated outside of transition relevance places (TRPs) (Sacks et al., 1974). In overlaps that are initiated within the TRP speakers use non-prosodic features for turn competition instead. For this reason we assume that f0 accommodation might be relevant only in overlaps that start before the TRP. Table 8 shows the performance of the Decision Tree (DT) classifier for all f0 features in overlaps that are initiated in these two contexts: at a TRP (atTRP) and before a TRP is reached (beforeTRP).

The DT results confirm our assumption. The duration of the overlap is significantly related to its competitiveness in both turn contexts, at TRP and before it is reached. The classifier based on f0 features, however, achieves an agreement of $\kappa = 0.15$ with human annotations for the classification of overlaps that are initiated within the TRP space (atTRP). This performance is not significantly different from that of the naive majority baseline classifier, which classifies all overlaps into the majority class of non-competitive and achieves the agreement with the human data of $\kappa = 0.00$. The performance is however significantly lower than that of the f0 based classifiers for overlaps that are initiated before a TRP ($\kappa = 0.25$). The f0 based classifier for overlaps with onsets before a TRP significantly outperforms the majority baseline. Therefore, in agreement with Kurtić et al. (2013) we can conclude that f0 accommodation features have only a small contribution to

Table 8

Decision tree classifier performance (Cohen's kappa scores) in two different turn contexts: within the TRP space (atTRP) and before a TRP has been reached (beforeTRP).

Feature group	atTRP	beforeTRP
f0	0.15	0.25 [*]
Duration	0.36 [*]	0.33 [*]

^{*} Indicates a significantly higher agreement with human categories than the naive majority classifier, which classifies all instances as the majority class (non-competitive for overlaps at TRP and competitive for overlaps before TRP).

the competitiveness classification for overlaps within the TRP space.

We further analyse the set of 495 overlaps (291 competitive and 204 non-competitive) that start before a TRP has been reached.

4.2. Competitiveness classification: f0 before TRP

For the set of 495 overlap instances that are initiated before a TRP, Table 9 shows the performance of classifiers trained and tested on normalising and initialising f0 features respectively. All feature sets classify turn competitive and non-competitive overlaps significantly better than the majority baseline. The classifier based on normalising features achieves the highest performance ($kappa = 0.28$) and outperforms the classifier based on initialising features ($kappa = 0.21$). However, this difference is not statistically significant.

Table 10 shows the normalising and initialising f0 features in different overlap contexts, i.e. in overlap (In) and immediately before its onset (Bef). The best performance is achieved by the classifier that uses overlapper’s in-overlap f0 features (InER) for competitiveness classification. The performance of all other classifiers however, is not significantly different from this classifier.⁷ Therefore we can conclude that all feature sets in Table 9 are relevant, although to different degrees, for designing a turn medial overlap as competitive or not.

Furthermore, to investigate the relative contributions of single features from the normalising and initialising feature sets (Table 9) and their combinations to competitiveness classification, we inspect the final DT model built from the entire data set with all features. The way in which these single features interact in the DT model is shown in Fig. 3.

It is known that DT structures can change fully or partly with changes in data sets even while their semantic stability is preserved (Last, Maimon, & Minkov, 2002; Landwehr, Hall, & Frank, 2005; Ntoutsis, Kalousis, & Theodoridis, 2008). We estimated the stability of our DT model by comparison of its performance to that of DT based ensemble classifiers, which are known to stabilise the variance of single DT models (Zhou, 2012). As noted above, we did not find a large discrepancy in performance between ensembles and single DT, which suggests the stability of the DT model. However, we additionally inspected DTs created for random stratified subsets output in each of the 10 cross-validation folds and found that not all segments of the DT model in Fig. 3 are equally stable. All 10 models share the top feature *Overlap types* and the first split. In all 10 models the recognitional overlaps are categorised as non-competitive. Also the order of features, with normalising features being placed above the initialising ones in the DT models is found in all 10 DT model versions.

The differences in models for random data subsets lie in the features used in sub-trees. The topology of the model categorising turn incursive overlaps is more stable than that for progressional overlaps. This is most likely due to the small data set size for progressional overlaps as discussed below.

Table 11 shows the versions of the sub-tree for turn-incursive overlaps that deviate from the general model

Table 9

Decision tree classifier performance (Cohen’s kappa scores) for normalising and initialising f0 features in overlaps that start before a TRP. The best performing classifier is in bold. The performance of all classifiers is significantly higher than majority baseline.

f0 features	Feature type	DT performance
All features	Normalising & Initialising	0.25
	Normalising	0.28
	Initialising	0.21

Table 10

Decision tree classifier performance (Cohen’s kappa scores) for normalising and initialising f0 features for overlaps that start before a TRP. The features are listed for each overlap context (i.e. in overlap (In) and just before overlap onset (Bef)) and speaker, i.e. overlapper (ER), overlappee (EE), as well as their combinations (see Section 3.3.2). The best performing classifier is in bold.

Feature type	Context and Speaker	DT performance
Normalising	BefEE	0.25
	InEE	0.24
	InER	0.31
Initialising	SimScore	0.26
	InIn	0.25
	InErBef	0.25
	InEeBef	0.24

(Fig. 3). In all DT models the same top feature $InER_{slope}$ is found. The initialising feature $InEeBef_{f0max}$ is also present in all models but one (submodel 3) and is placed below the normalising f0 slope feature. The variability relative to the general model is present in pre-overlap f0 height ($BefEE_{height}$) and the difference in standard deviation ($InErBef_{f0stdev}$).

Progressional overlaps exhibit the largest variation in model topology. Alternative topologies for progressional overlaps are shown in Table 12. The main similarity of the models to the all data model is that slope of overlappee’s pre-overlap talk ($BefEE_{slope}$) is placed at the top nod for progressional overlaps in 8 out of 10 models. The remaining 2 progressional overlap models have $InER_{slope}$ as the top feature. The most striking difference to the all data model is the presence of the initialising features in the models built on subsets. The feature $SimScore_{SelfEE}$ that captures the similarity in overlappee’s talk in and before overlap, for example occurs in half of all models for progressional overlaps. Furthermore, the between-speakers initialising features $InIn_{slope}$ and $InErBef_{slope}$ are found in 3 of the subset models. They indicate the difference in slope between speakers in overlap and to the TCU just before.

From these results we can draw the following conclusions about the relevance of f0 accommodation in turn competition:

All DT models indicate that the use of f0 accommodation for turn competitiveness depends on the type of the overlap onset initiated before a TRP (cf. Section 3.2.2). In this turn context we find three overlap types:

Recognitional overlaps are placed upon the recipient’s achieving the point of “semantic adequacy” (Jefferson, 1983c), i.e. at the point where a recipient has enough information about what the current speaker is going to say in the rest of the turn, but before a TRP has been reached. In Jefferson’s terms, recognitional means “placing one’s talk upon recognition rather than upon imminent completion of the recognised object.” (Jefferson, 1983c, p. 15). Jefferson (1983c) also notes that “These, often deeply turn incursive, onsets can be seen to be *principalled and reasonable*.” This characterisation of

⁷ As indicated by the two-tailed paired *t*-test, at the significance level of $p < 0.007$, with Bonferroni correction for 7-fold multiple testing.

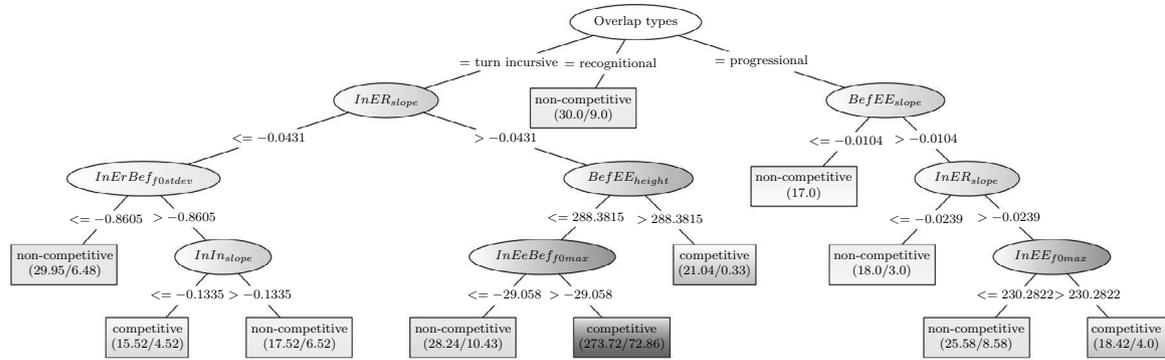
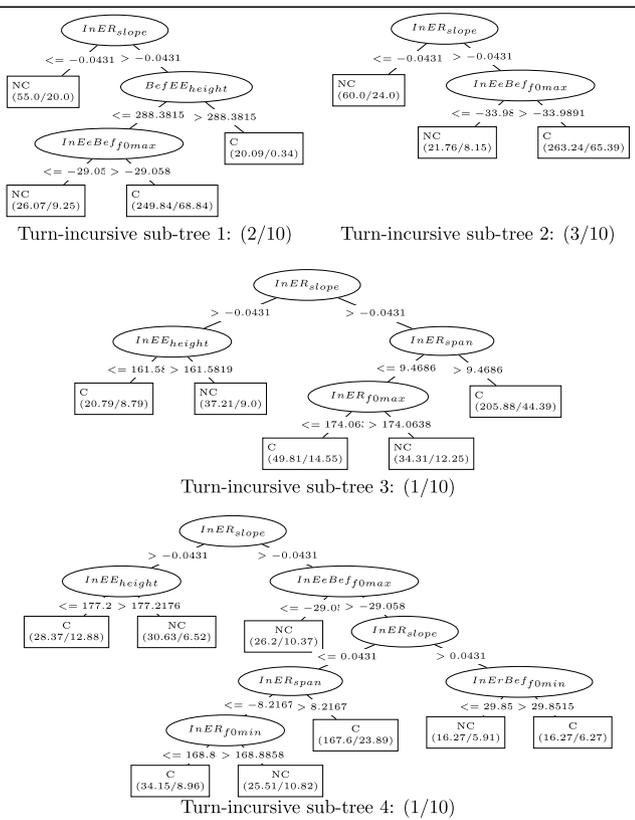


Fig. 3. Decision Tree model for overlap competitiveness classification built for the entire data set. The ellipses represent the features according to which the data set was split. The boxes represent the leaves of the tree and denote the categories into which the instances were classified (“competitive” (red) and “non-competitive” (green)). The first number is the weight of instances reaching the leaf (represented by the saturation of the opposite colour). Therefore, red colour also appears in a “non-competitive” box and green colour appears in a “competitive” box. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

Table 11
Decision tree model variants for random data subsets: Sub-trees for **turn-incursive overlaps**. C denotes competitive overlaps, NC non-competitive overlaps. The figures below the trees (e.g. 2/10) show how many out of 10 subset models share this topology. If not shown here a subset model has the same topology as the main model. 3 turn incursive sub-trees share the main model structure.



recognitional overlaps suggest that they are non-competitive. If so, it would be consistent with our data, where 21 out of 30 recognitional overlaps have been agreed upon by both annotators as non-competitive and only 9 as competitive (cf. Table 6). All DT models classify all 30 recognitional overlaps as non-competitive, with 9 competitive overlaps misclassified⁸.

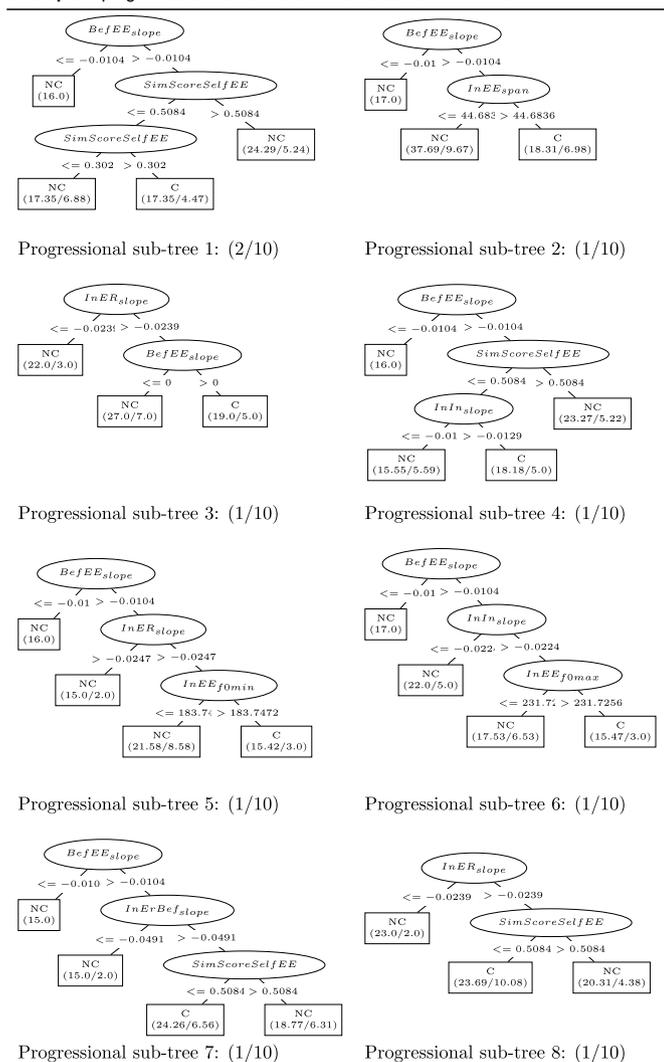
⁸ The weight (9.0) is given instead of the total number of misclassified instances because there is missing data in the data set. See Witten et al. (2015, pp. 212–213).

Progressional overlaps are overlaps initiated upon overlappee’s disfluency in the ongoing turn (Jefferson, 1983a) and often involve a collaborative completion of the turn in overlap (Local, 2005; Lerner, 2004). These overlaps are generally deployed and treated by the speakers as non-competitive (French & Local, 1983; Wells & Macfarlane, 1998). However, the current speaker’s disfluency might also offer an opportunity for the overlapper to take over the turn (Local, 2005; Lerner, 2004). In our data set 26 out of 79 (32.9%) progressional overlaps have been annotated by both annotators as turn competitive. This is a small data set for DT modelling, so all conclusions we draw about progressional overlaps need to remain provisional and need replication with more data.

As expected for this size of the data set the DT models indicate that the prosodic design of progressional overlaps is variable. The model for the entire data set in Fig. 3 suggests that progressional overlaps are realised by normalising f0 features. The most prominent normalising feature is the slope of overlappee’s f0 before overlap (*BefEEslope*, in 8 out of 10 subset models and the main model). However, the results for data subsets (Table 12) suggest that initialising features also might be relevant. They are found in 6 out of 10 models. The most prominent initialising feature is the *SimScoreSelfEE*, i.e. the similarity of overlappee’s in-overlap talk to that preceding the overlap. The models show that non-competitive overlaps are characterised by high (> 0.51) or low (≤ 0.30) intra-speaker similarity score. Furthermore, between-speaker initialisation is also found in progressional overlaps. A low difference in slope between speakers in overlap (*InIn_slope*) and just before (*InErBef_slope*) classify overlaps as non-competitive.

Turn incursive overlaps are overlaps initiated clearly before a TRP has been reached, but cannot be classified as either recognitional or progressional. There is little explanation in the literature of the possible conversational functions of these overlaps. Particularly puzzling are turn incursive non-competitive overlaps. Given that these overlaps could have been placed in the TRP space, a position where non-competitiveness is more likely, the question arises, why overlappers invade the ongoing turn if no turn competition is planned. 130 out of 386 (33.7%) turn incursive overlaps in our data set are marked by both annotators as

Table 12
Decision tree model variants for random data subsets: Sub-trees for **progressional overlaps**. 1 progressional sub-tree shares the main model structure.



non-competitive. The DT models suggest that speakers use both initialising and normalising f0 accommodation to realise a turn inceptive overlap as competitive or not. The best split of the data set into competitive and non-competitive turn inceptive overlaps is achieved by the normalising feature capturing the in-overlap f0 slope of the overlapper's talk ($InER_{slope}$) – a result stable across all DT models.

The non-competitive turn inceptive overlaps are generally characterised by a falling in-overlap slope (cf. 'slope' in Table 7). However, for some portions of data competitive overlaps are better modelled by an additional larger difference in standard deviation values between the overlapper's in-overlap talk and the overlappee's pre-overlap talk ($InErBef_{ostdev}$) in combination with a low in-overlap f0 slope difference between the speakers ($InIn_{slope}$).

4.3. Summary

Our results suggest that f0 accommodation is a relevant resource for turn competition in overlaps that are initiated before a TRP and that this is not the case for overlaps with

onsets within the TRP space. The investigation of the overlaps initiated before a TRP shows that normalising and initialising f0 features are both relevant for marking turn competition or its absence in this turn context.

The decision tree models for the entire data set and its random subsets reveal that f0 accommodation in turn competitive incomings differs according to the conversational function of the overlap. While recognitional overlaps were non-competitive regardless of their f0 design, f0 accommodation features were relevant to classify progressional and turn inceptive overlaps as competitive and non-competitive. In both overlap types, it was normalising features, the overlappers' and overlappees' slope of f0, that achieve better purity of split of overlaps into competitiveness categories and are positioned higher in the tree. The normalising f0 slope features are best predictors of competitiveness. The initialising features split the trees below the normalising f0 slope features and contribute to a more precise division of overlaps pre-classified by normalised features. The models selected difference based initialising features for turn inceptive overlaps. For progressional overlaps the intra-speaker f0 adaptation measure *SimScore* was most frequently used in subset models although the evidence about the design of progressional overlaps is weak given the variability of model topologies in the small data set.

5. Discussion

In this paper we investigated f0 accommodation as a turn competitive resource in overlap. Our study can be seen as exploratory in that it comprises an intensive analysis of a single naturally occurring conversation. Therefore the generalisability of our findings is subject to further investigations with more data and other conversation types.

Experiment outcomes. We studied four possible ways in which f0 accommodation may be deployed as a turn competitive resource (cf. Section 1). The above results indicate that all our findings only apply to the overlaps that are initiated before a transition relevance place has been reached, suggesting that f0 accommodation is only relevant for turn competition in these turn contexts (Table 8).

First, our results confirm that f0 initialisation, i.e. f0 accommodation in the local context of turn containing an overlap, is relevant for turn competition (Point (1), Section 1). As Table 9 shows, the DT model based on initialising f0 features classifies overlap as turn competitive and non-competitive significantly better than the majority baseline, which does not use any f0 information.

Second, we aimed to examine the relationship between f0 initialisation and f0 normalisation, i.e. the accommodation in f0 to speakers overall conversational norms without reference to local overlap context. To this end either f0 initialisation or f0 normalisation can be used for turn competition (Points 2 and 3 in Section 1). We found that normalising f0 features distinguish better between competitive and non-competitive overlaps than initialising features, however, this difference was not statistically significant (Table 9 and Section 4.2). Therefore we conclude that neither f0 normalisation nor f0 initialisation alone is used for turn competition in overlap.

This suggests that both f0 initialisation and f0 normalisation take place when speakers compete for the turn in overlap

(Point 4, Section 1). The combination of the two feature sets significantly outperforms the majority baseline which does not use any f0 information at all (Table 9). However, the performance of combined initialising and normalising features is not a significantly better discriminator than either of the feature sets individually. To understand the relationship between two types of accommodation features an inspection of the DT models was necessary. We analysed both the DT model built on the entire data set (Fig. 3) and 10 models built on random subsets of data. This revealed that different f0 accommodation features are deployed according to the different conversational functions of overlap. Progressional overlaps exhibited a lot of variance in their prosodic design due to the size of the data set and only allow tentative conclusions. Half of all models used normalising features only, while in the remaining DT models, initialising features were present. The most prominent initialising feature in progressional overlaps was *SimScore* indicating overlappee's adaptation in f0 to her pre-overlap talk. In turn incursive overlaps between-speaker f0 differentiation features were relevant. In models of both progressional and turn incursive overlaps normalising features achieve higher purity of split and are placed higher in the tree than the initialising ones. The initialising features make a more precise division of overlaps pre-classified by normalised features. This suggests that for turn competition in overlap normalisation constrains the local f0 accommodation. However, a more thorough investigation of statistical interdependencies between normalising and initialising f0 accommodation measures is necessary to be able to finally assess how these feature groups interact in relation to turn-competition in overlap.

Prosody and interactional meaning. These results on f0 accommodation as a turn competitive resource offer evidence from naturally occurring talk for describing the nature of the relationship between prosodic constructions and their interactional meanings. The relevant question here is: Are prosodic constructions mapped to their meanings before being employed in conversation, or do they emerge in a case-by-case fashion as conversation participants co-create them in the course of a conversation? Three possible scenarios describe this form-function relationship:

1. If participants compete for the turn, they draw on pre-existing fixed individual patterns independently of the prosodic design of the current turn and independently of each other.
2. If participants compete for the turn, they will do so by adapting to or differentiating from what the co-participant or they themselves do at the given moment. There are no stored prosodic patterns and all prosodic behaviours are co-constructed in a case-by-case manner.
3. If participants compete for the turn they use fixed patterns that do turn competition in overlap, but they are further accommodated in the local overlap context.

Our results indicate that both normalising and initialising features are relevant turn competitive resources. Furthermore, the inspection of DT models revealed that initialising features select overlaps pre-classified by normalising features. Therefore, our exploration allows the conclusion that there are fixed patterns that participants use, but that these are accommodated in the local turn context both between speakers and relative to the same speaker's pre-overlap design.

Contributions to relevant previous work. In relation to previous work, these results make several contributions. First, they provide evidence that expands the current state of knowledge on the role of f0 accommodation for turn competition in overlap. So far only f0 normalisation has been investigated as a turn competitive resource (French & Local, 1983; Kurtic et al., 2009; Kurtić et al., 2013; Lee et al., 2010). In this work we added the initialising view on f0 accommodation and tested its relationship to f0 normalisation with respect to turn competition in overlap. Therefore this work fills the gap in the existing work on prosody of turn competition in overlap and contributes to the broader enterprise of interactional phonetics and related disciplines that are concerned with how different prosodic patterns are systematically used and oriented to by conversation participants in order to convey interactional meanings (Couper-Kuhlen & Selting, 2006).

The results presented here also contribute to the large and growing body of work on prosodic accommodation (cf. Section 2) which generally does not investigate instances of overlapping talk.

Finally, this work adds to the discussion of possible approaches to studying f0 accommodation in conversation recently opened by Zellers and Schweitzer (2017). They describe and evaluate the initialising and normalising approach to interpreting pitch accommodation (referred to as pitch matching in their paper). After evaluation of both approaches in Zellers and Schweitzer (2017) their results remain inconclusive as to which one is preferred to studying prosodic accommodation in conversation. Therefore Zellers and Schweitzer (2017) conclude that it might be the task at hand that should determine which approach to take. Our study considered both these approaches for the "task" of turn competition in overlap. Our results go along with those in Zellers and Schweitzer (2017) who find that there is evidence in the literature for both views of prosodic accommodation, which would indicate that prosodic patterns indeed carry recognisable interactional meanings, but in some overlap instances their use might need to be additionally constrained by a between-speaker accommodation of pitch features. This means that at least until further evidence is provided any study on prosody in conversation needs to consider both normalising and initialising prosodic features.

6. Conclusions

This paper investigates f0 accommodation as a turn competitive resource in overlapping talk. We assess two approaches: initialisation (i.e. local accommodation to f0 design of overlapped turns) and normalisation (accommodation to participants' general norms without reference to local context). Previous work offers evidence for normalisation as a turn competitive resource. However, this is the first study on f0 initialisation and the relationship between both approaches in relation to turn competition in overlap.

We found that f0 accommodation is only relevant as a turn competitive resource in overlaps that start clearly before a speaker transition becomes possible. In this turn context, we found that normalising and initialising f0 features can both be relevant turn competitive resources. Their deployment depends

on the conversational function of overlap (recognitional, progressional or turn incursive). While speakers do not use f0 to differentiate between competitive and non-competitive recognitional overlaps, they use both normalising and initialising f0 features in progressional and turn incursive overlaps. The relationship between normalising and initialising features is hierarchical, with normalising features pre-classifying the overlaps into competitiveness classes and initialising features providing more specific division of these classes.

These findings imply that interactional meaning in conversation is achieved by general prosodic patterns mutually interpretable by speakers that are additionally accommodated to prosodic events in the local context.

This work contributes to the literature on prosodic resources for turn competition in overlaps, to the studies of prosodic accommodation and to the question on whether normalisation or initialisation better describes prosodic accommodation in relation to interactional meaning in naturally occurring talk. The investigation presented here is an intensive analysis of a single naturally occurring conversation. Therefore the generalisability of our findings is subject to further investigations with more data and other conversation types.

Future work. Future work could extend our methodology to be able to investigate f0 contours in the context of turn competition more precisely. It is also necessary to adapt the feature extraction algorithms to be able to look beyond “f0 contours”, e.g. intensity, where it is necessary to take into account the volatile fluctuations in signal energy due to articulation processes on the phone level. Future research should also improve extraction techniques to better cope with aperiodic phonation phenomena, as they appear e.g. in creaky voice. A focus on glottal pulses rather than fundamental frequency (as a periodic phenomenon) may help at this point.

Concerning the calculation of pitch contour similarity, one future improvement could be to question the domain on which the calculations are done. Instead of looking at entire stretches of speech, as we did with respect to turn-constructive units delimited by TRPs, one could follow the traffic-light model developed by Wells and Stackhouse (2015). The key focus lies there on the tonic syllable(s) that could serve as domain for analysing contour adaptation or differentiation. Furthermore, the in-overlap development of turn competition in the manner of Schegloff (2000) could be studied to better understand overlap resolution.

Finally, a previously unattempted methodological novelty would be to scale up to large sets of data the analysis of non-conforming (deviant) cases as it is currently practised in Conversation Analysis in a qualitative example-by-example manner. This would allow information contained in deviant cases to be exploited in a methodological approach like ours that depends on mining for patterns in large data sets of naturally occurring talk.

Acknowledgements

This work was supported by the *Agence Nationale de la Recherche*, Grant No. ANR-12-JCJC-JSH2-006-01, *Arts and Humanities Research Council*, Grant No. AH/L009307/1 and *Leverhulme Trust*, Grant No. EM-2015-017. This research has been performed while Emina Kurtić was affiliated with the Department of Human Com-

munication Sciences, University of Sheffield. We would like to thank Bill Wells for continuous discussions on the topic and proof reading the manuscript, Ella Page for help with the annotation campaign. And thanks to the Institute of the German Language for allowing Jan Gorisch to continue this research.

Appendix A. Examples of competitive and non-competitive overlap

This section shows examples of competitive and non-competitive overlap, as well as an unclear case, where annotators disagree on overlap competitiveness. It illustrates the process of sequence analysis used to arrive at overlap competitiveness decisions.

Competitive. An example of an overlap annotated as clearly turn competitive by both annotators is given in Fig. 4.^{9,10}

The overlap instance of interest here is at the beginning of line 2. It is embedded in the conversation where the participants speak about an upcoming stag do of a friend. Speaker m002 is about to relate a story about his cousins having attended a stag do in the past (line 1). He aims to describe their behaviour during the stag do they attended. He first introduces the story giving the background information on his cousins' usual behaviour (*my cousins argue all the time like.*) by which he secures his turn space to complete the story. Then he makes the main point (*And they were nearly like fighting in middle of the street.*), which is overlapped competitively by f003.

f003 comes in with *The girls are going at a point where neither the entire story, nor m002's ongoing turn And they were nearly like [fighting in middle of street.] are syntactically or semantically complete. After the overlap region, m002's topic of arguing and fighting cousins is interrupted in favour of the conversation on girls' activities from line 2 onward. The overlapper f003 has secured her turn by overlapping m002. She even has the opportunity to wait until the end of f002's "clearing throat" (l.2) before changing the focus from cousins to girls. f003 asks the question *Whose stag do did they go on last weekend.*, which is answered by f001 after a 0.8 s silence (*Girls can't go on stag dos though*) and possibly also by the interrupted speaker m002 (Oh). The changes in speakership and conversation topic at the point of incompleteness classify this overlap in line 2 as competitive in terms of Definition (3.1). The annotators have identified this in their unanimous competitiveness rating of 5.*

Non-competitive. An example of an overlap annotated as clearly non-competitive by both annotators is given in Fig. 5 in line 4 where speaker f003 overlaps with the last bit of speaker f001's turn *I thought you said (.) do you know what their company specialised.* The participants talk about f003's father's who used to own a bike shop in the past. f001 is interested in the kind of bikes that were sold in that

⁹ The example has been created with the Partitur-Editor in EXMARaLDA (Schmidt, 2001), a tool for segmentation, transcription and visualisation of talk. The conversation is presented here in the musical score format, where the instruments are the speakers. Their utterances and non-verbal behaviour are presented from left to right and in line blocks from top to bottom, with overlap instances evident from the time-alignment in the transcript. The overlaps are shown as parallel text within a line-block.

¹⁰ The full stop at the end of a stretch of an utterance indicates the end of a Turn Constructive Unit (TCU). It does *not* signify any prosodic properties, e.g. a final pitch fall, as they are used in other transcription conventions such as HIAT (Rehbein, Schmidt, Meyer, Watzke, & Herkenrath, 2004, Section 4.2).

[1]	m002	But they were just like my cousins argue all the time like. And they were nearly like
[2]	m002	fighting in middle of street.
	f002	((clearing throat))
	f003	The girls are going. Whose stag do (.) did they go on last
[3]	f001	Girls can't go on guys' g* stag dos though.
	m002	Oh.
	f002	Hen dos.
	f003	weekend. (0.8) No no but the girls the hen do
[4]	f001	Oh.
	m002	No it were-
	f002	Lu sh. (0.1)
	f003	they're havi* the hen do's in marbella. (0.8) And the bo*- I know.

Fig. 4. Transcript of a sequence containing a clearly competitive overlap in line [2].

[1]	f001	What like rally bikes. Racing
	f003	But he used to own a (0.3) bike shop. (1.2) (0.7) Racing bikes.
[2]	f001	ing bikes. Oh.
	f003	And mountain bikes. Specialise d. Do you know the company Specialised.
[3]	f001	What- Oh. ((laughter)) I thought you said (.)
	f003	(0.2) (0.8)((yeah)). (0.6) It's a bike company.
[4]	f001	do you know what their company specialise d. ((laughter)) I was like- ((0.6s laughter))
	f002	Do they.
	f003	No no.
[5]	f001	((0.7s laughter)) No. ((inbreath)) ((inbreath)) Oh no. I don't really know anything about
	m002	((0.7s laughter))
	f003	((0.7s laughter))
[6]	f001	bikes. ((clearing throat)) (1.8)
	f003	Yeah. (0.2) I wonder if i'm gonna be dog sitting. (0.9)
[7]	f002	Yeah. That'd be fun.

Fig. 5. Transcript of a sequence containing a clearly non-competitive overlap in line [4].

shop. f003 specifies them (Racing bikes. And mountain bikes.) and mentions a bike brand: Specialized (mid of line 2), immediately followed by the question Do you know the company Specialized. f001 indicates that she does not understand: What- and the conversational flow is interrupted with pauses, a very soft word that sounds like yeah.

f003 clarifies that It's a bike company (l. 3) showing that she treats the interruption in conversational flow as f001's ignorance about Specialised being a brand name of a bike company. Getting to know where her misunderstanding came from, f001 laughs and spells out what she thought of having

heard: do you know what their company specialised instead of do you know the company specialised. f003's No no in line 4 confirms that there was a misunderstanding that is now clarified. f001 laughs during that overlap and starts to clarify her puzzlement I was like- that gets overlapped by a third speaker who takes on the misunderstanding by jokingly inverting it and taking the misunderstood question seriously: Do they as in: does the company specialise in some types of bikes? This provokes further laughter (also by other speakers) until speaker f001 closes the sequence with Oh no. I don't really know anything about bikes

(ll. 5–6). The end of the sequence is marked by a long silence of 0.9 s and f003's change in topic to dog sitting (ll. 6–7).

The overlap *No no* in line 4 was not designed to take over the floor and was not treated as such by the prior speaker f001. An agreement and a common understanding were achieved, and the overlap was used to confirm this. The completion of the sequence without an evidence of disruption to either speaker's turns in this overlap indicates that this overlap is non-competitive in the sense of Definition 3.1. The annotators reflect this in their unanimous competitiveness rating of 1 (clearly non-competitive).

Unclear. The example in Fig. 6 shows an overlap (line 3) whose competitiveness is disagreed upon by the annotators. The example starts with a short discussion about the current room temperature (line 1–2). However, this is only briefly picked up by the interlocutors and is summarised by the same speaker who brought it up (f002: ((2.1s)) *Think it's just the pressure*). After this speaker f001 introduces a new topic starting with a strong inbreath and *I think i might apply for coach trip next year*. The topic is applying for a reality TV show called *Coach Trip*.

In line 3 the overlapper m002 comes into f001's turn at the point where f001 has supplied just enough information for the interlocutors to understand her plans: Action (*apply*), What for (*coach trip*) and When (*next*), assuming that it is clear to everyone that this TV-show is produced on a yearly basis. This is treated by m002 as an upcoming potential completion of her turn, and he starts his turn (*I'm*), which is a legitimate thing to do when another speaker has come to a TRP (Sacks et al., 1974). But f001 does not stop at this TRP. Instead, she produces an add-on to her turn, *Be like oh i've got nothing to do*. This turn that may supply a reason for applying (she has nothing to do) or might be analysed as a case of face-saving self-mockery action (Zare, 2016) as there is no evidence

that going on reality TV is generally accepted and f001 does not know how the others will evaluate her plans. This may be the reason why despite being in overlap, she insists on completing this add-on. m002 positively evaluates and reinforces this plan, which is evident from his claim that he is *definitely gonna apply for it* as opposed to the overlappee f001's more tentative *I think i might apply...* Since this supporting action is in overlap with f001's add-on, m002 reiterates the overlapped portion *Definitely gonna apply for it* in a delayed completion (cf. Lerner, 2002) in line 4. f001 acknowledges his support for the idea and the willingness to join in by suggesting to *go together* in the following turn and another add-on *Would be amazing* (ll. 4–5). This is acknowledged with laughter from all participants. The overlap instance in line 3 displays both competitive and non-competitive elements.

Arguments for 'competitive'. In this overlap both speakers may be seen to compete for the turn: m002 continues his turn and does not acknowledge that f001 is producing an add-on *Be like oh i've got nothing to do*, although she shows that she does not want to give up the floor. m002 is clearly aware that he is talking in overlap because he produces a delayed completion *Definitely gonna apply for it*. (in ll. 4–5). Delayed completions (Lerner, 1989) are not only a strategy for making the in-overlap talk heard, but also a competitive strategy for taking over the speaking turn in cases where it is not clear from the in-overlap talk who will win the turn.

f001 continues her add-on despite the possibility to drop out upon realisation that m002 treated her prior turn as complete and started a new turn at a legitimate place.

The annotator who identified this overlap as competitive indeed marked that both speakers are competitors in this case.

Arguments for 'non-competitive'. On the other hand, this overlap can also be seen as non-competitive: The overlapper m002 initiates the overlap at a legitimate point.

[1]	m002	I'm quite a good tempera
	f002	*s hot in here. (0.8) Anyone else hot. Or is it just me. (0.1) Mm.
	f003	Just you.
[2]	f001	((inbreath)) I think i might apply for coach trip next
	m002	ture.
	f002	(2.1) Think it's just the pressure.
[3]	f001	year. Be like o::h i've got no thing to do. ((laughter))
	m002	I'm definitely gonna apply for it.
	f003	((laughed) Yeah.)
[4]	f001	((inbreath)) We sh*d totally go together. Would be a
	m002	Definitely gonna apply for it. It was a-
[5]	f001	mazing. ((laughter))
	m002	((laughter)) We'd probably break every single re* probably stay on for all fifty
	f002	((laughter))
	f003	((laughter))
[6]	m002	days to be honest.

Fig. 6. Transcript of a sequence containing an unclear case of overlap in line [3].

The overlappee f001 does not (see it as necessary to) place her overlapped talk outside of the overlap, but with We sh*d totally go together orients to m002, acknowledging his right to the turn in the preceding utterance Definitely gonna apply for it. At the end of the sequence, both speakers have completed their turns and arrived at a mutual agreement. Therefore, the sequence can also be interpreted as closing without evidence of interruption to either party. These arguments allow to classify the overlap as non-competitive according to Definition 3.1.

In addition, the actions in this segment are collaborative, which may bias the annotator towards marking the overlap as non-competitive. f001's announcement to go on a coach trip is met with support and reinforcement by the overlapper m002. There is also a lot of laughter by all participants, which evidences collaboration and agreement, rather than opposition in action.

The complexity of the sequence is reflected in the annotators' disagreement on the overlap's competitiveness. However, while the annotator who identified the overlap as competitive was confident of the analysis of this overlap as clearly competitive (5), with both parties competing, the other annotator gave it a rather non-competitive rating of (2).

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. In *Proceedings of the international conference on language resources and evaluation (LREC)* (pp. 3051–3111).
- Aimetti, G. A. P. (2011). *A computational model of early language acquisition: A data-driven approach inspired by the empiricist view of cognitive development* Ph.D. thesis. University of Sheffield.
- Bonin, F., De Looze, C., Ghosh, S., Gilmartin, E., Vogel, C., Polychroniou, A., Salamin, H., ..., & Campbell, N. (2013). Investigating fine temporal dynamics of prosodic and lexical accommodation. In *Proceedings of interspeech*.
- Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, 96, 41–44.
- Bunt, H. (2011). Multifunctionality in dialogue. *Computer Speech & Language*, 25(2), 222–245.
- Burgoon, J. K., Stern, L. A., & Dillman, L. (2007). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge University Press.
- Calhoun, S., & Schweitzer, A. (2012). Can intonation contours be lexicalised? Implications for discourse meanings. *Prosody and Meaning (Trends in Linguistics)*, 25, 271–327.
- Çetin, Ö., & Shriberg, E. (2006). Analysis of overlaps in meetings by dialog factors, hot spots, speakers, and collection site: Insights for automatic speech recognition. In *Ninth international conference on spoken language processing*.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893.
- Chowdhury, S. A., Danieli, M., & Riccardi, G. (2015). Annotating and categorizing competition in overlap speech. In *Proceedings of ICASSP 2015* (pp. 5316–5320).
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cooke, M. (1993). *Modelling auditory processing and organisation*. Cambridge University Press.
- Core, M. G., & Allen, J. (1997). Coding dialogs with the DAMSL annotation scheme. In *AAAI fall symposium on communicative action in humans and machines* (Vol. 56). Boston, MA.
- Couper-Kuhlen, E. (1996). The prosody of repetition: On quoting and mimicry. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: Interactional studies* (pp. 366–405). Cambridge University Press.
- Couper-Kuhlen, E., & Selting, M. (2006). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: Interactional studies* (pp. 11–56). Cambridge University Press.
- Cruttenden, A. (1997). *Intonation*. Cambridge University Press.
- De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4), 1917–1930.
- De Looze, C., Oertel, C., Rauzy, S., & Campbell, N. (2011). Measuring dynamics of mimicry by means of prosodic cues in conversational speech. In *Proceedings of the 17th international congress of phonetic sciences (ICPhS), Hong Kong* (pp. 1294–1297).
- De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, 58, 11–34.
- French, P., & Local, J. (1983). Turn-competitive incomings. *Journal of Pragmatics*, 7(1), 17–38.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, N. Coupland, & J. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics*. Routledge communication series (pp. 1–68). USA: Psychology Press.
- Gorisch, J., Wells, B., & Brown, G. J. (2012). Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech*, 55(1), 57–76.
- Hedberg, N., Sosa, J. M., Görgülü, E., & Mameni, M. (2010). Prosody and pragmatics of Wh-Interrogatives. In *Proceedings of the 2010 annual meeting of the canadian linguistics association*.
- Heldner, M., Edlund, J., & Hirschberg, J. (2010). Pitch similarity in the vicinity of backchannels. In *Proceedings of interspeech* (pp. 3054–3057).
- Hutchby, I., & Wooffitt, R. (2008). *Conversation analysis*. Polity Press.
- Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: A classification perspective*. Cambridge University Press.
- Jefferson, G. (1983a). Another failed hypothesis: Pitch/loudness as relevant to overlap resolution. *Tilburg Papers in Language and Literature*, 38.
- Jefferson, G. (1983b). Another failed hypothesis: Pitch/loudness as relevant to overlap resolution. *Tilburg Papers in Language and Literature*, 1(38), 1–24.
- Jefferson, G. (1983c). Two explorations of the organisation of overlapping talk in conversation, 1: Notes on some orderliness of overlap onset. *Tilburg Papers in Language and Literature*, 1(28).
- Jefferson, G. (1986). Notes on 'latency' in overlap onset. *Human Studies*, 9(2), 153–183.
- Kelly, J., & Local, J. (1989). *Doing phonology: Observing, recording, interpreting*. Manchester University Press.
- Kousidis, S., Dorran, D., McDonnell, C., & Coyle, E. (2009). Time series analysis of acoustic feature convergence in human dialogues. In *Digital media centre conference papers*. Dublin Institute of Technology.
- Kousidis, S., Dorran, D., Wang, Y., Vaughan, B., Cullen, C., Campbell, D., ... Coyle, E. (2008). Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues. In *Proceedings of interspeech 2008*. Dublin Institute of Technology.
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411–433.
- Kurtić, E. (2011). *Overlapping talk and turn competition in multi-party conversations* Ph. D. thesis. University of Sheffield.
- Kurtić, E., Brown, G. J., & Wells, B. (2009). Fundamental frequency height as a resource for the management of overlap in talk-in-interaction. In D. Barth-Weingarten, N. Dehé, & A. Wichmann (Eds.), *Where prosody meets pragmatics (Studies in Pragmatics 8)* (pp. 183–204). Emerald Group Publishing Limited.
- Kurtić, E., Brown, G. J., & Wells, B. (2013). Resources for turn competition in overlapping talk. *Speech Communication*, 55(5), 721–743.
- Kurtić, E., Wells, B., Brown, G. J., Kempton, T., & Aker, A. (2012). A corpus of spontaneous multi-party conversation in bosnian serbo-croatian and british english. In *Proceedings of the international conference on language resources and evaluation (LREC)* (pp. 1323–1327).
- Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14(4), 334–339.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, 59(1–2), 161–205.
- Last, M., Maimon, O., & Minkov, E. (2002). Improving stability of decision trees. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(02), 145–159.
- Lee, C.-C., Black, M., Katsamanis, A., Lammert, A. C., Baucom, B. R., Christensen, A., ..., Narayanan, S. S. (2010). Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proceedings of interspeech 2010* (pp. 793–796).
- Lerner, G. (1989). Notes on overlap management in conversation: The case of delayed completion. *Western Journal of Speech and Communication*, 167–177.
- Lerner, G. H. (2002). Turn-sharing: The choral co-production of talk-in-interaction. In C. Ford, B. Fox, & S. Thompson (Eds.), *The language of turn and sequence* (pp. 225–256). Oxford University Press.
- Lerner, G. H. (2004). Collaborative turn sequences. *Pragmatics and Beyond New Series*, 125, 225–256.
- Levinson, S. C. (1983). *Pragmatics (cambridge textbooks in linguistics)*.
- Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies: Short papers-volume 2* (pp. 113–117). Association for Computational Linguistics.
- Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of interspeech*.
- Local, J. (2003). Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics*, 31(3–4), 321–339.
- Local, J. (2005). On the interactional and phonetic design of collaborative completions. *A figure of speech: A festschrift for John Laver*, 263–282.
- Matarazzo, J. D., Weitman, M., Saslow, G., & Wiens, A. N. (1963). Interviewer influence on durations of interviewee speech. *Journal of Memory and Language*, 1(6), 451.
- Müller, F. E. (1996). Affiliating and disaffiliating with continuers: Prosodic aspects of reciprocity. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: Interactional studies* (pp. 131–176). Cambridge University Press.

- Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5), 790.
- Ntoutsis, I., Kalousis, A., & Theodoridis, Y. (2008). A general framework for estimating similarity of datasets and decision trees: Exploring semantic similarity of decision trees. In *Proceedings of the 2008 SIAM international conference on data mining* (pp. 810–821). SIAM.
- Oertel, C., Włodarczyk, M., Tarasov, A., Campbell, N., & Wagner, P. (2012). Context cues for classification of competitive and collaborative overlaps. In *Proceedings of speech prosody 2012*.
- Ogden, R. (2006). Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, 1752–1775.
- Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observers. *Journal of Nonverbal Behavior*, 30(4), 157.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190.
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2–3), 203–228.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers.
- Rehbein, J., Schmidt, T., Meyer, B., Watzke, F., & Herkenrath, A. (2004). Handbuch für das computergestützte Transkribieren nach HIAT. *Arbeiten zur Mehrsprachigkeit, Folge B*, 56, 1. ff.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 696–735.
- Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. *Studies in Interactional Sociolinguistics*, 13, 52–133.
- Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1), 1–63.
- Schmidt, T. (2001). The transcription system EXMARaLDA: An application of the annotation graph formalism as the basis of a database of multilingual spoken discourse. In S. Bird, P. Buneman, & M. Liberman (Eds.), *Proceedings of the IRCS workshop on linguistic databases, 11–13 December 2001* (pp. 219–227). Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania.
- Selling, M. (2010). Affectivity in conversational storytelling: An analysis of display of anger or indignation in complaint stories. *Pragmatics*, 20(2), 229–277.
- Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., & Carvey, H. (2004). *The ICSI meeting recorder dialog act (MRDA) corpus* Technical report. Berkeley, CA: International Computer Science Inst.
- Shriberg, E., & Stolcke, A. (2004). Direct modeling of prosody: An overview of applications in automatic speech processing. *Speech Prosody, 2004*, 575–582.
- Sicoli, M. A., Stivers, T., Enfield, N. J., & Levinson, S. C. (2015). Marked initial pitch in questions signals marked communicative function. *Language and Speech*, 58(2), 204–223.
- Szczepek Reed, B. (2006). *Prosodic orientation in english conversation*. Palgrave MacMillan.
- Szczepek Reed, B. (2012). Beyond the particular: Prosody and the coordination of actions. *Language and Speech*, 55(1), 12–33.
- Ten Have, P. (2007). *Doing conversation analysis*. Sage.
- Truong, K. P. (2013). Classification of cooperative and competitive overlaps in speech using cues from the context, overlapper, and overlappee. In *Proceedings of interspeech 2013* (pp. 1404–1408). International Speech Communication Association.
- Wells, B., & Macfarlane, S. (1998). Prosody as an interactional resource: Turn-projection and overlap. *Language and Speech*, 41(3–4), 265–294.
- Wells, B., & Stackhouse, J. (2015). *Children's intonation: A framework for practice and research*. John Wiley & Sons.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2015). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Zare, J. (2016). Self-mockery: A study of persian multi-party interactions. *Text & Talk*, 36(6), 789–812.
- Zellers, M., & Ogden, R. (2014). Exploring interactional features with prosodic patterns. *Language and Speech*, 57(3), 285–309.
- Zellers, M., & Schweitzer, A. (2017). An investigation of pitch matching across adjacent turns in a corpus of spontaneous German. In *Proceedings of Interspeech* (pp. 2336–2340).
- Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. CRC Press.