Using prosodic cues to identify dialogue acts: methodological challenges

1. Introduction

As part of a larger project to identify communicative functions in corpora and to seek potential speech technological applications of interactional research, we were interested in ways to automate parts of the analysis of speech data without considering its lexical content. Associating prosodic features with particular conversational events implies that, if these features are operationalised appropriately (note that a lot hangs on what counts as "appropriate" here), it should be possible to identify the associated conversational events. At the very least one should be able to test the extent to which prosodic features do, in fact, correlate with discourse uses.

Previous research was encouraging in this regard. Many studies within the field of interactional linguistics have shown that prosodic features such as intonation can play a key role in the way interactions are shaped and monitored by participants (e.g. Couper-Kuhlen and Selting 1996; Selting 1996; Local 2004; Barth-Weingarthen et al. 2009; Szczepek Reed 2009, 2010, 2015). The main implication of these authors' research findings is that words, sequential context and prosody (as well as other potential resources) all interact to shape meaning in interaction. Certainly, humans do not normally interact in situations where prosody/phonetics are separated from lexicon, grammar, sequential context or the other locally co-constructed aspects of interaction (Szczepek Reed 2011). Still, there may be circumstances, in a particular sequential context, where prosody alone may *distinguish* one social action, or dialogue act, from another.

Competitive overlaps emerged as a good candidate dialogue act: it had previously been found that phonetic features distinguish competition from non-competition in overlapping talk (see

1

e.g. French and Local 1983; Wells and Macfarlane 1998; Schegloff 2000; Kurtic et al. 2009, 2013; Kurtic 2011). In this paper we test the extent to which such phonetic features are independent from lexical/syntactic content, with an overarching aim to operationalise dialogue features in a way that may be exploited at a larger scale. The applicability of findings from conversation analytic and interactional linguistic research has not often been tested in a speech technological context, which is what we set out to do. This can then serve as an alternative, or addition, to existing speech technological tools.

In Section 1.1 we summarise the phonetic features found to correlate with turn competition, and we also introduce an illustrative example. Then, in Section 2.1, we introduce the methods used to automatically extract overlaps from the audiofiles. In Section 2.2 we describe competitive and non-competitive overlaps as defined in the manual annotation, and in Section 2.3 we present the phonetic parameters used in the analysis. We present the results in Section 3, starting with the overall differences between competitive and non-competitive overlaps (Section 3.1), followed by a decision tree analysis that further tests the extent to which prosodic features can distinguish the two types of overlap (Section 3.2). In Section 4 we account for exceptions to the overall findings based on qualitative analyses, and in Section 5 we discuss our findings and the methodological challenges involved.

1.1 Phonetic features in competitive overlaps

The first systematic study on phonetic correlates of competitive overlaps was conducted by French and Local (1983), who found that speakers use phonetic features to distinguish whether or not they are competing with a co-participant's turn trajectory (overlapper), and whether or not turn-production is yielded to the co-participant (overlappee). They found that overlappers tend to use higher pitch and intensity (tagged <H + F>) in competitive than in non-competitive speech, and that overlappees either yield the floor by stopping their speech,

or return the competition by increasing speech intensity without changing pitch (<F> but not <H>). They concluded that <H + F> is the only cue which is uniquely indicative of the competitiveness of the overlap, and argued against the role of pragmatic function and lexical design in turn competition.

Wells and Macfarlane (1998) also found that <H + F> is an important cue. However, they argued that competitiveness in overlaps is not independent from its placement relative to the co-participant's turn. They found that competitive overlaps are more likely to occur prior to the nuclear pitch accent of a turn than after the nuclear pitch accent (however, not all pre-nuclear overlaps are competitive). Kurtic et al. (2013) supported this view, and even found that placement of overlap played a greater role than prosodic features in indicating competitiveness. Other features have been suggested as well, including speech rate (Schegloff 2000), but the relevance of this feature in competitiveness is disputed (Kurtic 2011). Kurtic (2011) also distinguished between a range of features and sub-features, related to pitch, intensity, duration, pausing, and speech rate. They confirmed that pitch and intensity (mean and range) offer important cues. Furthermore, Kurtic (2010/2011) suggested that, although weak, intra-overlap pausing and speech rate are significant cues to competitiveness in overlap (see Kurtic 2011 for a comprehensive review of the organisation of overlap in relation to turns of talk; see also Zuraidah and Knowles 2006; Oertel et al. 2012).

Example 1 is illustrative of competitive overlap. In this example, the caller (C) cannot find a file he currently works on. C claims to have transferred the file (via FTP) from his lab computer (referred to as the hard-drive in line 02), in order to access the file from his home computer. The competitive overlap occurs in lines 14-17. The transcription is a simplified version of the GAT2 transcription system (Selting et al. 2011; see Appendix for more detail).

```
Example 1

01 A:       I think you should go BACK to THAT(p)? (.) (b) (0.4)
02          that maCHINE and see if it's on the HARDdrive there.
03          (0.5)
04 C:       (yeah/see) I TRIED that and I? I cʰ every o? every
05          time I FOUND it it was the wr?(0.4) the one that
06          was SHORTer.
07          (0.7)
08 C:       that didn't have all the reVISions.
09          (0.7)
10 C:       .MHH=
11 A:       =.pthh
12          (1.9)
13 A:       and did? (0.2)
14    ->    are you SU[RE that you CALL[ED it the [same=
15 C: ->             [and the:        [the?       [the o?
16 A: ->    =[FI [LE   name.  ]
17 C: ->    =[(.)[the ONly rea]son like I'm really WORRied
18          about it is like FOUR? there  was like (.)
19          (tn) (0.2) NINE pages (uh) (.) compared to FOUR pages
20          (this is like)- (0.4) .mhh (0.4) quite [a lot of] WORK.=
21 A:                                             [mm      ]
22 A:       =.hhh .tk and(m)? (1.0) okay .hh and uhm (1.8) so
23          you're SURE you saved under the same FILE name.
24          (0.2)
25 A:       it's not [a new (d)] (.) different FILE.
26 C:                [yeah.    ]
27 C:       yeah.
```

In lines 04-06, C rejects the possibility suggested by the Agent (A) (lines 01-02) that the latest version of C's file might still be available on his lab computer, as C has already checked this. Following line 06, there are multiple opportunities for A to take up on C's rejection (lines 07, 09, and 11-12), but A remains silent until his inbreath in line 12, followed by a verbal response in lines 13-14. A's response is treated as problematic for the progress towards a solution to C's problem. In particular, C builds on his previous turn with a relative clause in 08: "that didn't have all the reVISions.". In general, such increments increase the relevance of a turn-transition as they do not contribute anything new to the interaction (Walker 2004). Here C makes an effort to ensure that A and C are indeed referring to the same file. Thus, within the larger interactional process, it is evident that the participants monitor each other's localised speech productions for content which may have led to trouble in understanding. With this in mind, notice that C also self-repaired his reference to the same

file in line 05, by stopping the production of the "wr?" (the "?" represents a glottal stop; see Appendix). A likely target here is "the wrong one/file", but C refers instead to "the one that was SHORTer" (lines 05-06).

C's efforts do not remedy A's problem by providing a candidate solution, as evidenced by the following sequence of pauses and inbreaths from A. The Agent eventually formulates a response (identifying a potential gap in C's argument), but not without problems: First A aborts his turn initiation in line 13 ("and did- (0.2)"); then, as A restarts the turn ("are you sure"), C overlaps with A's turn (line 15) and they both proceed in overlap.

In lines 13-20, A and C initiate two quite different trajectories and uptakes on the preceding talk: while A pursues the possibility that C might have saved the file under a new name/location, C initiates an account explaining why it is so problematic for him not to have found the file. For C, it is not only that the (old) file has been revised and is shorter than the latest version, but it is also much shorter than the latest version , and thus represents a significant amount of labour (the latest version is 9 pages whereas the old version is only 4 pages long). By providing this account C attempts to demonstrate the seriousness of his problem, treating the preceding silences as a potential threat to the validity of his call.

We can regard the overlap in lines 14-17 as competitive as the participants compete for the projection of different conversational trajectories, as opposed to collaborating on the same conversational outcome. In phonetic terms, while A does not yield the floor to C, C increases loudness and pitch of his speech, from each speech incoming to the next, until he speaks in the clear. Figure 1 captures C's overlapping speech (blue square) represented by pitch (middle section) and loudness (green line in bottom section). During the overlap, A returns the competition by maintaining his intensity; however, he does not increase loudness.[1] These

---

[1] It is important to keep in mind that the absence of observable orientations from co-conversationalists is not

characteristics are in line with the findings in previous research as described earlier in Section 1.[2]

In what follows we test the extent to which increased pitch and loudness, and some other phonetic features, may be distinctive of turn-competitive overlaps, compared to non-competitive overlaps.
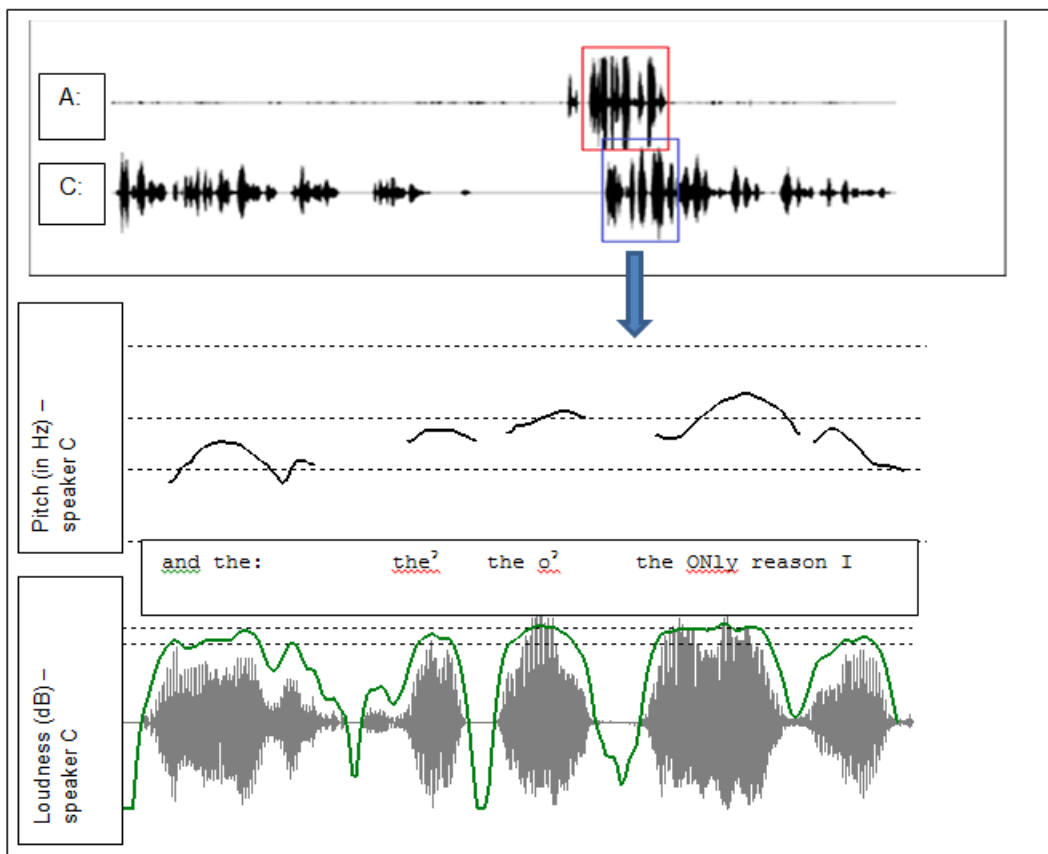

Figure 1. Waveform of turns 04-15 in Example 1 (top: Agent in channel 1 and Caller in channel 2). Coloured frames highlight the competitive overlaps in lines 12-15. Caller's speech is provided with movement in pitch (middle) and intensity (bottom).

---

necessarily counter-evidence against a given social practice: i.e. participants do not need to behaviourally orient to rules for them to be relevant to them (Robinson 2007: 71). Much successful communication is achieved implicitly, but it is accountable-for if challenged.

[2] It might be relevant at this stage to provide a contrastive example, i.e. of non-competitive overlap. Due to space limitations, however, we refer to previous publications for such examples (see e.g. Kurtic 2011).

2. Material and Methods

The dataset used was collected from a UK based call-centre. There were in total 335 calls in this corpus, 38 (11%) of which were randomly selected for further analysis. We labelled all overlaps in these 38 calls, based on a combination of case-by-case interactional analysis, distinguishing competitive and non-competitive overlaps. The data were recorded in stereo.

2.1 Automated extraction of overlaps

The identification of overlaps to be analysed was automated, based on a speech/non-speech detection algorithm. On a technical level we refer to speech elements between two non-speech elements as "speech chunks" (see e.g. Edlund and Heldner 2005 for similar terminology). The overlaps were defined on the basis of temporal attributes of individual speech chunks themselves, and the temporal relationships between speech chunks in the two recording channels. The definition rules of overlaps is summarised in Figure 2, with time on the horizontal axis, and speech (speaker A top, speaker B bottom) represented as chunks in the clear and in overlap.
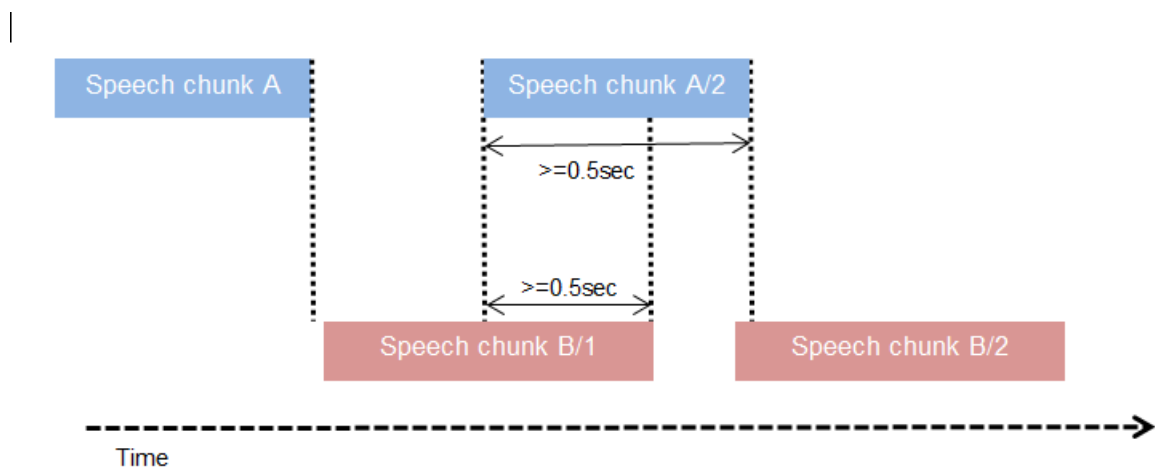

Figure 2. Overlapping turns operationalised. The simultaneous speech chunks A/2 and B/1 would be regarded as overlaps in this case. The shorter (terminal) overlap A/2 and B/2 would not be regarded as overlap with our definitions ('terminal overlap' – see exception rule #1).

The rules and exceptions are as follows:

- The individual speech chunk (A) is 0.5 seconds or longer, and the start of Speech chunk A overlaps with speech in the other channel (speech chunk B)

- Exception #1: If speech chunk A starts less than 0.5 seconds prior to the end of speech chunk B and speech chunk A ends later than 1 second after the end of speech chunk B, speech chunk A is not counted as overlap (these are largely terminal overlaps – see Jefferson 1984)

- Exception #2: If speech chunk A is shorter than 1 second, and starts/ends in the proximity of a brief non-speech chunk in B, it is not counted as overlap (these are largely continuers – e.g. "mhm", "yeah")

Note that our definition of overlaps is entirely based on acoustic and durational features: for analytic clarity we have built our analysis independently of lexico-syntactic content at this stage of the analysis. Although participants might be sensitive to the timing of overlaps relative to pitch accents (see Wells and MacFarlane 1998) rather than a specific time boundary, the minimum boundaries we set for overlaps seemed consistent with the exclusion of terminal overlaps (Jefferson 1984) and continuers. While the automatic detection of phrasal prominence was not a feasible part of this study, it would be worth exploring in the future.

We included all overlaps that met the requirements listed above regardless of quality of sound. By comparison, Kurtic removed groups of overlaps from the analysis due to poor quality because they were "not sufficient for reliable automated analysis" (2011: 55). There are obvious arguments in favour of Kurtic's approach; however, our decision to include all physical overlaps was made with a view to our future goal to be able to deal with speech corpora which have not been pre-treated by humans for linguistic labelling.

2.2 Manual annotation of overlap types

The operationally defined instances of physical overlap (as summarised in Setion 2.1) were then labelled by analysing them separately and in the context of surrounding talk. Competitive and non-competitive overlap were the two main annotation categories used. The labelling was informed by the conversation analytic (CA) method. CA is a case-by-case analysis attending to participants' own orientations to talk as it emerges, rather than analysts' *a priori* interpretations of what is happening (Sidnell 2011). While CA research traditionally excludes the large-scale and quantitative analyses typical of corpus linguistic research (Robinson 2007; Tracy 2007), recent research shows how quantitative methods can be successfully combined with a CA approach (Stivers et al. 2009; Stivers and Enfield 2010; Kurtic 2011; see also Steensig amd Heinemann 2015 for an overview and discussion).

In line with our bottom-up, CA-inspired approach to the analysis, only instances where the overlap is observably used as a resource for butting in to the current trajectory of talk (see Example 1 above), and/or treated by the co-participant as disruptive to their on-going talk, was labeled as competitive. Our definition of competitive overlap is as follows:

> *Instances of physical overlap where there is observable interactional evidence that either one or both speakers seek to gain speakership at the expense of the co-participant's talk.*

Our definition is similar to Kurtic's (2011), but unlike Kurtic (2011) we included as instances of competitive overlap single words which, at least momentarily, seek to claim the floor, at a point in time when the co-participant is also talking. These were labelled competitive as long as there was sufficient interactional evidence that the co-participant would (or in fact did) treat them as disruptive, e.g. by recycling their speech. Lexical and pragmatic content of the

overlapping turn was used as part of our annotation decisions, but secondarily to the observable use/treatment of it in the interactional context. In order to test the extent to which phonetic features distinguish our (and participants') interactional categories, we avoided making any labelling decisions based on loudness, pitch, duration, etc.

Using the above definition, we labelled overlaps as non-competitive when there was no strong evidence of competitiveness. Previously, non-competitive overlaps have been regarded as "licensed overlap" (Schegloff 2000), and are comprised of response tokens (Schegloff 1982), terminal overlaps (Jefferson 1984), collaborative completions and choral productions (see Lerner 1996).

Not all overlaps detected were actual speech, and some were caused by background speech (e.g. within the call centre). Therefore we also used a third main annotation category: noise. Sub-categories to each main label were also defined (we do not use these in this paper which is deliberately coarse grained).

2.3 Phonetic parameters

The phonetic parameters considered included average pitch, pitch range, average intensity, intensity range and durations. Each measure was performed for both overlapper and overlappee channels.

Whether or not a speaker increases their pitch and loudness depends on how loud or high-pitch their previous normal speech is. Therefore, while measuring pitch and intensity as absolute values, we also calculated the *relative* increase in pitch and loudness, with reference to the speaker's own speech in the recording. This is what we refer to as normalisation. Two types of normalisation procedures were used for pitch and intensity. The first used the value in a current overlap (e.g. A/2 in Figure 2 above) with reference to the average value for all

non-overlapping within-speaker talk in the relevant recording (as done by Kurtic 2011). In a second procedure we normalised the value of a current overlap with reference to that of the preceding (non-overlapping) speech chunk (e.g. A/2 with reference to A/1 in Figure 2).

Duration was defined on the basis of a speech/non-speech detection algorithm. Intonation was defined as pitch range based on its slope/error in the speech signal. We chose a robust linear model in order to deal with spurious values. All averages were based on the median. The left boundary was defined as the acoustic starting point of the overlap. The measure included 0.2 sec following the physical overlap for the overlapped chunk (if this exceeded the right boundary of the overlapping chunk). This is in line with Oertel et al. (2012) who found that features of turn competition may continue slightly beyond the acoustic endpoint of the overlap. Unlike Kurtic (2011), we did not include speech rate (i.e. frequency of syllables) in our analysis, nor the intensity range. This is not to deny or ignore the potential relevance of these features, but due to what operationalisation procedures we currently have available and choose to focus on.

3. Results

We hypothesised that competitive overlaps (Competitive Turn Incursions; henceforth CTI) would have higher (normalised) pitch and loudness values, greater pitch and loudness range, and longer durations than non-competitive overlaps (Non-Competitive Turn Incursions; henceforth NCTI). Section 3.1 summarises the average overall differences between CTIs and NCTIs, and Section 3.2 explores how different features might work together in characterising CTIs as opposed to NCTIs.

3.1 Average differences between competitive and non-competitive overlaps

Our data contain a total of 323 overlapping chunks, 246 labelled NCTI, while 42 were labelled CTI.[3] Thus there is a majority of non-competitive overlaps, including utterances such as "yeah", "okay", but also more explicit terms such as "that's great, thank you".

In terms of phonetic features, there are some average differences for duration and pitch range values between NCTI and CTI overlaps, which all go in the expected direction of higher pitch and intensity values in CTI compared to NCTI (see Figure 3). Another pattern that emerges in Figure 3 is that the overlapped channel (OD) tends to have higher pitch values than the overlapping (OP) channel, whereas the opposite relationship appears to be the case for intensity – OP has the highest values. This finding seems somewhat in disagreement with French and Local (1983), who found that only overlapping speakers seem to raise pitch in competitive overlaps, not overlapped speakers.
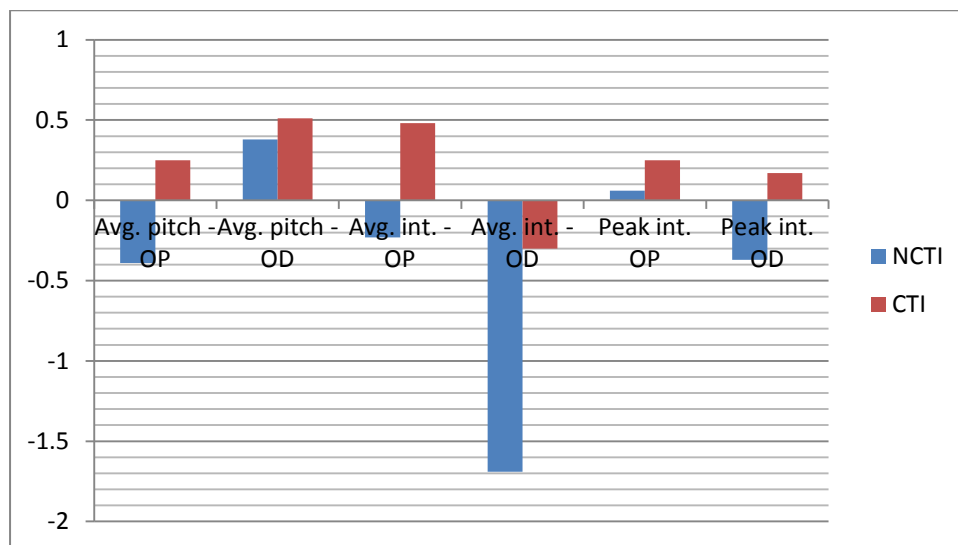

Figure 3. Normalised pitch and intensity values for overlapping (OP) and overlapped (OD) chunks, for non-competitive (NCTI) and competitive (CTI) overlaps. Pitch values are given in average (median) and intensity in average and peak values.

Are these phonetic differences alone sufficient to distinguish competitive and non-competitive overlaps? The answer is no, and this becomes more evident when examining the plots below for normalised average pitch values and normalised average intensity values

---

[3] Forty chunks were labelled non-speech "Noise" and are excluded from this paper.

(Figure 4). As we can see, although the average values are slightly higher in CTIs than NCTIs, the values overlap. None of the average differences are statistically significant using Wilcoxon signed-rank test.
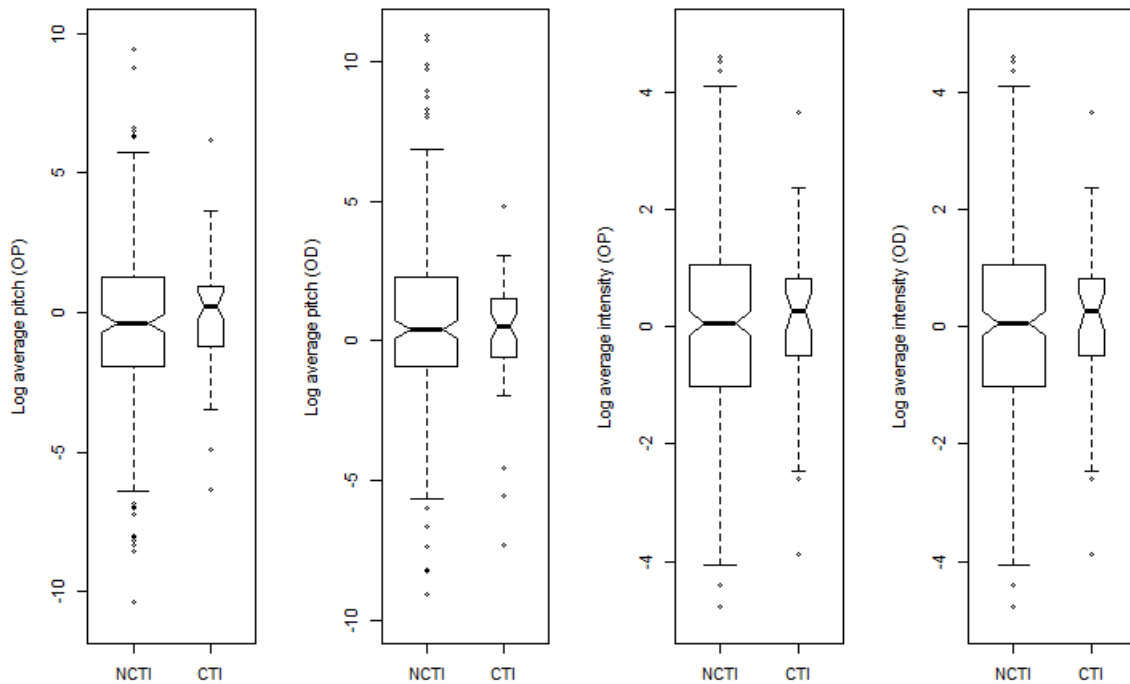

Figure 4. Boxplots of average (median) normalised pitch and intensity values, for overlapping (OP) and overlapped (OD) chunks, in non-competitive (NCTI) and competitive (CTI) overlaps. Box sizes are determined by number of observations.

The normalised values reported so far are based on overall values for each speaker. The literature reports that the use of intensity and pitch in competitive speech can be observed when comparing the overlap with the previous speech by the same speaker (e.g. French and Local 1983; Oertel et al. 2012). We also normalised the values for overlaps against previous speech chunks (see Section 2.3), and found some tendencies for CTIs to have higher positive peak intensity difference relative to preceding speech, than NCTIs. The difference was largest

in the overlapped channels (Log peak intensity OD; see Figure 5), and this difference was

significant at 5% level using Wilcoxon signed-rank test (W = 6595, p-value = 0.004187).
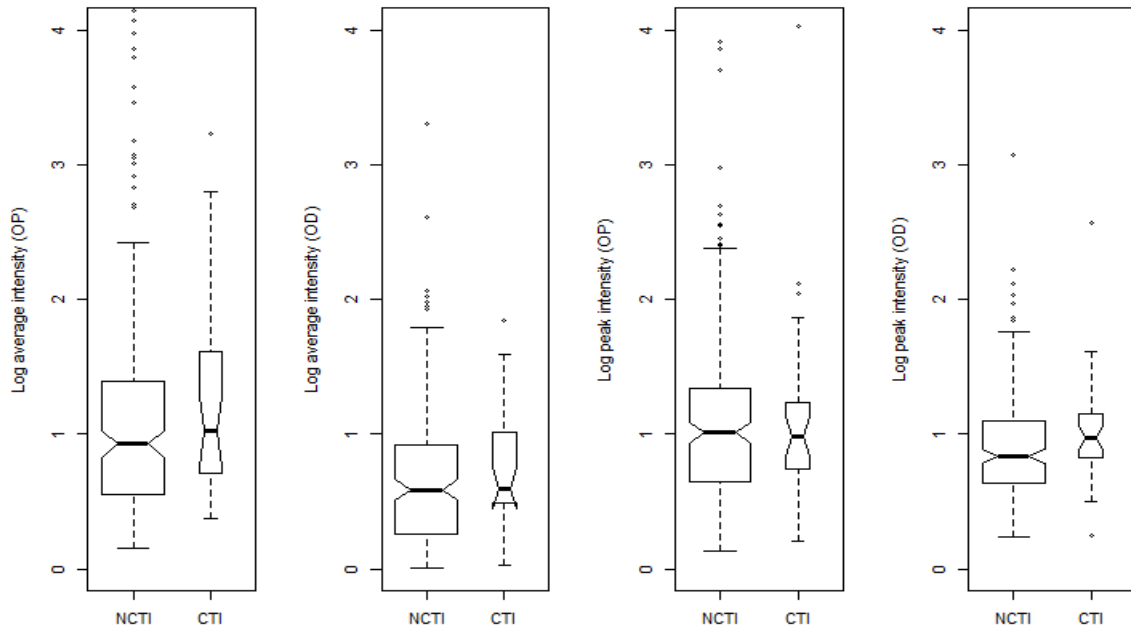

Figure 5. Boxplots of average (median) pitch and intensity values normalised according to
preceding turn, for overlapping (OP) and overlapped (OD) chunks, in non-competitive (NCTI)
and competitive (CTI) overlaps. Box sizes are determined by number of observations.

In summary, the overall results presented above do not suggest that phonetic features *alone*

may distinguish between competitive and non-competitive overlaps. This is confirmed with a

Principal Component Analysis (PCA). The PCA uses all the normalised pitch and intensity

features in our analysis, and convert them into linearly uncorrelated variables. The first 5

principal components (which together account for most of the variation) are plotted in Figure

6 below: if the phonetic cues alone were able to distinguish CTI/ NCTI, then we should see a

much clearer division between the three labelled groups than is apparent here.

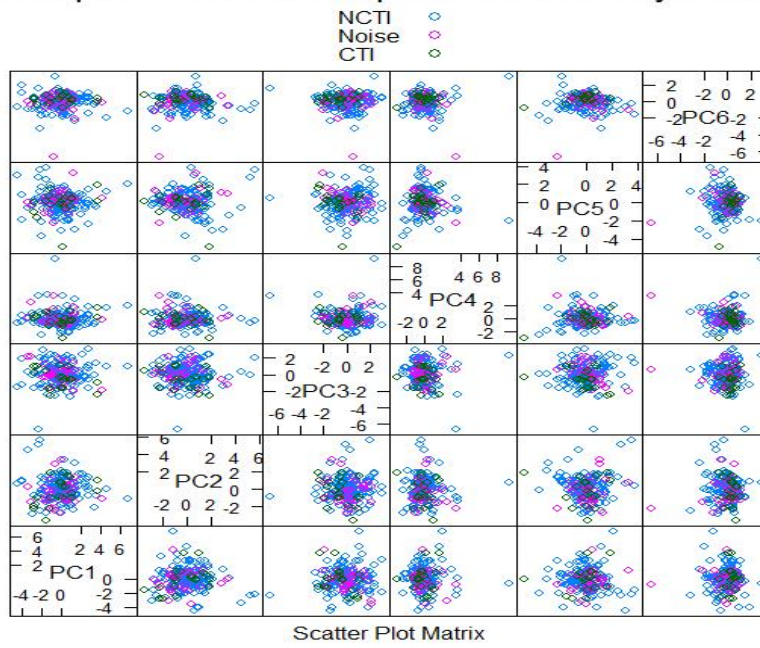Figure 6: XY plot for Principal Components, using all pitch and intensity values in our analysis.

3.2 Decision trees

The key to identify competitive overlaps without reference to lexical content may lie in how the different phonetic features work together. To investigate this further we used decision tree modelling. Decision trees are useful in exploring the relative contribution of the different variables. Based on decision tree modelling we can establish the extent to which competitive and non-competitive overlaps are distinguishable based on the features we measured (see Breiman et al. 1984)

One known issue with decision trees is that large differences in group size create a bias against potential group distinctions (Cieslak and Chawla 2008). Therefore we down-sampled the NCTI group to 42, equal to the size of CTI. To further address whether the parameters had any predictive value for the data, we performed a 10-fold cross-validation, which provided us with error scores to inform us of how well the model generalises to unseen data (Baayen 2008). Since the error score will increase with each added node in the decision tree, we pruned the tree to the size that best represented the predictive value of our data. The pruning method was determined based on conventions used in Baayen (2008). The pruned decision tree is shown in Figure 7, which shows that intensity values interact with pitch slope (our definition of intonation): in general it appears that CTIs have higher intensity values and less negative or flatter pitch contour than NCTIs.
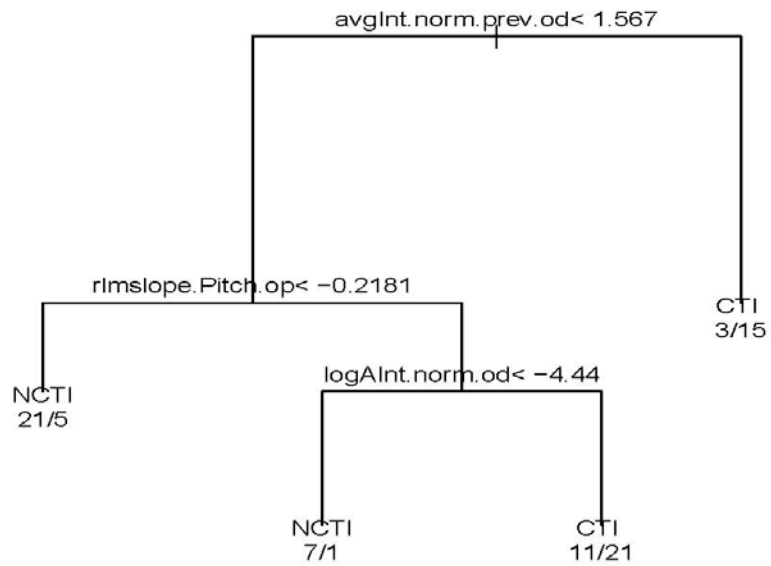
Figure 7: Pruned Decision tree, showing three key features to the identification of competitive overlaps: (i) average intensity normalised according to previous turn in the overlapped channel (avgInt.norm.prev.od), (ii) pitch slope in overlapping channel (rlmslope.Pitch.op), and (iii) average intensity values in the overlapped channel (logAInt.norm.od). The decision tree shows how many CTI/NCTI overlaps fall under and above the given values.

We note that when the intensity value is above a threshold (> 1.567) the overlap is probably competitive (estimated to occur in 15/18 cases). When the intensity value is below that threshold, and the pitch slope is sufficiently negative (< -0.218), the overlap is likely to be non-competitive (in 21/26 cases). When the pitch slope is less negative or positive, the overlap is likely to be non-competitive if the intensity value is very low (< -4.44), in 7/8 cases, otherwise the overlap is fairly likely to be competitive (21/32 cases).

These results correspond to previous findings, in terms of competitive overlaps tending to have higher intensity values than non-competitive overlaps. The relevance of pitch is perhaps less clear than has previously been found: the relevance of pitch to the distinction competitive and non-competitive overlaps is perhaps better defined by its slope (intonation) than its average.

To address how well our findings can predict whether a given overlap is competitive, we built a predictive model (based on the decision tree) that would allocate each overlap as CTI or NCTI. We found that 6/42 (14%) overlaps manually labelled as competitive were falsely predicted as NCTI. On the other hand, 14/42 (33%) manually labelled non-competitive overlaps were falsely predicted as CTI (see Table 1). If we combine these false allocations we get 23.8 percent ((6+14 false allocations)/84 overlaps) misclassifications in the data. In other words, if we use only phonetic features to classify the overlaps we would be wrong about one quarter of the time.

18

Table 1: Observed categories against Decision Tree choice. The table shows the proportion the model selects as CTI (CTIchoice = TRUE in rows 3-4) which is actually CTI (72%), and the proportion of NTCI (CTIchoice = FALSE in rows 1-2) which is actually NCTI (82.3%).

	observed	prediction	Freq	Prop. correct
1	NCTI	FALSE	28	82.3%
2	CTI	FALSE	6	
3	NCTI	TRUE	14	
4	CTI	TRUE	36	72%

4. Negative examples of competitive overlap

These results suggest there are many exceptions to previous claims about competitive speech, and a wide phonetic overlap between CTI and NCTI. Our research has identified a number of exceptions to the norm regarding phonetic features of competitiveness: that is, instances where putative phonetic features for competitiveness are used, but are neither designed, nor dealt with as being competitive. One would expect to find examples where a turn is designed to be competitive, but the co-participant does not observably deal with it as competitive (see Kurtic 2011). We have taken this further, and have identified examples of overlaps which both in their interactional placement and design as well as the manner in which they are oriented to, blur or possibly disprove that there is a clear-cut distinction between CTI and NCTI. We will present one such example associated with laughter; laughter has previously been noted as a regular location for overlap (e.g. Ford and Thompson 1996).

In Example 2 the Agent (Carla) and Caller (Bill) address an issue with access to a network server. Bill has suggested he might have caused a crash of the server, which was designed and treated as humoristic. In lines 01-02, Carla follows up with her story from the early days, when she managed to shut down her office's computer system ('spot') by typing the "the KILL command BACKwards,".

```
Example 2
01 Carla:   BACK in the EARly d(h)ays I did actually shut SPOT
02          down o[ne time .hhh [TY[Pi[ng th]e KILL=
03 Bill:        [HHH    hi   [hi[hi[hi   ]
04 Carla:   =[comm]and BACK[wards,=
05 Bill:    =[(hi)]         [.hh   =
06 Carla:   =oops hhuh [hh huh huh [huh]
07 Bill:               [oops   hhh [huh] huh huh=
08 Bill:    =.hh[hh    [oh:    no:    ]
09 Carla:       [.hh an[d the system a]dministrator ((...))
```

The "I did actually" in 01 projects the upcoming telling as a genuine event, with reference to a topic they so far have presented as a joke. Carla's story is thereby a highly legitimate

second story in this context (Sacks 1992). The first possible turn-completion in line 02 comes following "SPOT down", after which Bill initiates his laugh response ('Spot' is designed and treated as shared knowledge in this sequence). It is worth noting that although Carla is reporting on unfortunate events, the participants continue to deal with them humoristically. Bill's laughter in 03 seems to facilitate Carla's story further, and is also fitting with Carla's treatment of her own turn while speaking (she laughs during "d(h)ays" in line 01). Bill's laughter in 03 occurs in overlap with Carla's turn. Carla does not hold back from talking but continues and the loudness and pitch increases during the prominent syllables of "TYPing the KILL COMMand BACKwards," compared to her previous talk. This is illustrated in Figure 8, in which we focus on the loudness of C's talk (bottom panel; green line), which increases during B's overlapping laughter.
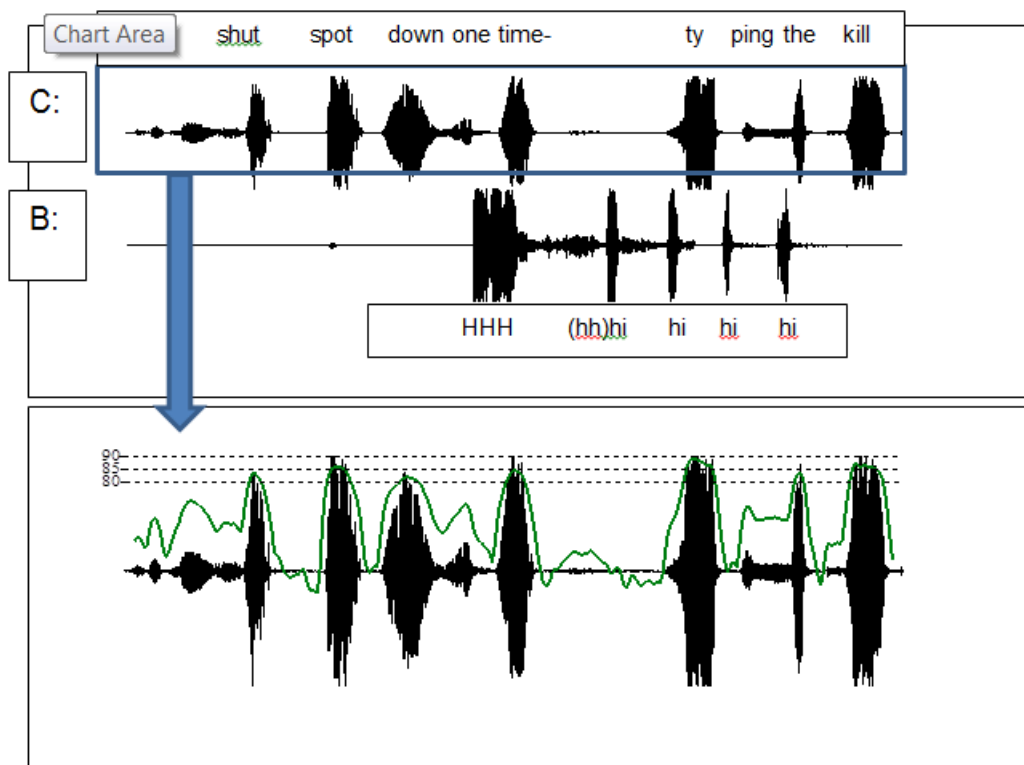

Figure 8. Waveform (bottom) showing the increased intensity in Carla's talk during overlap.

The increasing pitch and loudness in the overlapping talk is part of the speakers' commentary on what has just been said: the overlapped speaker initiates the commentary while the overlapping speaker affiliates with her commentary. Importantly, the fact that speakers increase pitch and intensity is not sufficient evidence for arguing in favour of competitiveness here.

There are several features here which support the notion of affiliation as opposed to competitiveness. First, laughter in itself is often used to display affiliation with a co-participant (Glenn 2003). Relatedly we note that the absence of laughter might have signalled disaffiliation here. Also, there is no compelling evidence for arguing that the Agent (Carla) treats the customer's incoming talk/laughter as competitive. Carla mirrors Bills's treatment of her turn ("oops" + laughter in lines 06-07). Prior to this Bill prepares his upcoming turn at the transition-relevance point at the end of line 05 with an inbreath. Before this Bill initiates laughter following possible transition-relevance place (TRP) in the early stages of line 05. These are used as collaborative devices, displaying anticipation of where A's turn is heading. Finally, in this case, the loudness feature seems to have more to do with the (shared) assessment action rather than the fact that two speakers are talking at the same time.

In summary, there is nothing in the interaction of Example 2 indicating that either speaker is competing for speakership, even though they use increased pitch and intensity. This example involves laughter; other counter-examples involve other forms of speaker engagement and collaboration. This leads us to some interesting questions: (i) do such counter-examples represent exceptions to the rule, and to what extent have these been addressed previously? (ii) If these examples indeed can be regarded as exceptions, how can we integrate them in our study, and more generally, with our understanding of overlap management? It might be that the negative examples represent quite specific sequential environments; but on the other hand, we might be dealing with more general phenomena, such as the display of (non-competitive)

enthusiasm and facilitative support, for which pitch and intensity are not necessarily used distinctively from competitive overlaps. For example, within a discourse-pragmatic perspective, Wichmann (2011: 189) argued that a dynamic use of pitch is important for the expression of strong emotion. The expression of strong emotion might be characteristic of both competitive (or disruptive) and non-competitive (non-disruptive) talk.

This also raises questions concerning the identification, and selection, of cases to be analysed: how are relevant cases of turn competition selected? How categorical are participants' own orientations to competitiveness vs. non-competitiveness, and thereby, how categorical can our analytic categories be? These questions pose problems both with regard to our own research, as well as for CA-based research more broadly. They highlight some of methodological challenges we are concerned with, which we discuss further in Section 5.

5. Discussion

This paper has presented our attempt to apply previous findings on prosodic/phonetic cues for dialogue functions to a large-scale corpus. We asked whether we could use phonetic cues to determine if an overlapping chunk of speech is competitive or non-competitive. Previous findings (French and Local 1983; Wells and Macfarlane 1998) suggest that competitive overlaps have higher intensity (overlappee) and higher intensity + pitch (overlapper) than normal for the speaker, and to a larger extent than in non-competitive overlaps. Two important reasons support this as a starting point in the quest to identify different dialogue acts based on prosodic features. Firstly, they were based on real-world interactions and the detailed analysis of participants' own orientations to dialogue functions. Secondly, the phonetic cues for competitiveness might be used independently of lexico-syntactic content (French and Local 1983).

As reported above, although competitive overlaps tend to have higher intensity (and flatter pitch/intonation) values than non-competitive overlaps, the results are weak in terms of discriminating the two overlap-types based on phonetic features alone, and overall the overlap categories are not clearly separated on the basis of the selected phonetic parameters. In an analysis of individual examples we indicated how, in interactional terms, loud and high-pitched overlaps do not necessarily correspond with competitive talk (i.e. talk which lacks overt display of collaborative co-ordination and/or affiliation between the participants). Thus, we conclude that competitive and non-competitive overlaps are not distinguishable on the basis of phonetic/prosodic cues alone.

Our findings appear contrary to previous work (French and Local 1983; Wells and Macfarlane 1998; Kurtic et al. 2009, 2010, 2013). Note, however, that we are not arguing that phonetic features such as intensity and pitch are never used by speakers to show that they compete for a turn. Indeed, our findings show an overall tendency for such features to be used in competitive speech. More to the point: loudness and pitch may in some circumstances feature as powerful resources for speakers to display competitiveness, in an attempt to "win" the turn at the cost of the co-participant. However, this does not imply that similar phonetic cues cannot in some circumstances be associated with non-competitive overlaps.

Apparently, then, there are differences in how certain activity types are associated with particular prosodic features: (i) with reference to their relevance to the participants themselves during actual interactions; and (ii) whether the proposed (and frequently observed) prosodic features can be used to define an activity type. At the core of this disparity, we argue, lies a methodological conundrum.

First of all, there is an obvious challenge regarding how to deal with the rich and complex nature of conversation in automated, speech analytic research. The complexity of talk is often

highlighted in research within interactional linguistics/CA. According to Local (2005: 264, reporting Laver 2003):

Talk-in-interaction relies on complex, highly structured, rule-governed behaviour of a semiotic richness that throws into sharp contrast the minimalist approach to the description of speech in spoken language favoured by recent generation of linguists and phoneticians.

On this view it is unsurprising that a few prosodic features like intensity or pitch cannot be used to classify dialogue functions. Speakers' comprehension of speech is experienced in real time, and involves a combination of many features simultaneously and sequentially organised, and knowledge, which cannot easily be reduced to a small subset (see also Szczepek Reed 2011). Clearly, prosody should be studied in connection to verbal and other elements in the construction of turns of talk. Whether verbal content would help us distinguish competitive and non-competitive overlaps however remains an open question.

From a quantitative perspective we find a conflict between adding variables to the analysis and testing which features are relevant for performing (and distinguishing) particular actions. Acknowledging the complexity of conversations is an important starting point for anyone interested in speech and language. However, operating simply is still a necessity at the level of analysis. For example, in order to study a particular interactional phenomenon it is necessary to operationalise categories of interactional events/actions and an associated set of features: sequential, phonetic, etc. This is also the case when the categories are informed by the participants' own orientations (the cornerstone of the CA approach). We would argue that attending to this conflict is an important research area for the future: progress within interactional linguistics may rest on how successfully we can create large-scale corpus analyses that are sensitive to interactional processes (see e.g. Deppermann 2010).

Other questions arising are how categorical we can expect prosodic and other linguistic cues to be. For example, turn competition is a relevant action category in the sense that participants may display orientations to turn competition using particular phonetic features. However, it is not clear that participants *necessarily* make such distinctions, because there is not always evidence in interactional data to suggest whether or not something is relevant to them (Robinson 2007). The distinction between deployment and treatment of actions has been highlighted in previous work on overlap (Wells and Macfarlane 1998; Kurtic 2011), where the co-participant's observable treatment of overlaps is central to the definitions of competitive overlaps. This raises questions regarding cases of non-observable reactions to the incoming overlap: does the absence of interactional treatment mean that there is no turn-competition? Philosophically, we are tempted by the suggestion implicit in the observation mentioned above, that if a participant does not react then it may be unclear whether a speaker was actually being competitive. For example, a speaker's intentions may only become accessible (interpretable) several turns after the fact (see Zeitlyn (forthcoming)). If treated as competitive it was competitive, but if not then its status may never be clearly established.

This takes us into heady and profound philosophical issues that we must leave for another place. But some much more practical issues also surface here. First of all: how do we deal with exceptions to a rule, and examples which are not a clear-cut member of any of the annotation categories? There are relevant questions to be raised regarding how relevant or interesting cases are selected for analysis, what is excluded and for what reasons. In our research we force-labelled overlaps as either competitive and non-competitive (or containing noise). Non-competitive overlaps included both discourse elements which may typically be quiet-spoken, e.g. backchannels/continuers ("yeah", "mhm"), as well as more facilitative responses ("oh really") and laughter, which may typically be produced louder and with higher pitch ranges than the speaker's average. There may be valid reasons for excluding such

overlap categories (see Kurtic 2011), and focusing only on typical listener signals (backchannels) may give a stronger separation between competitive and non-competitive overlaps. However, ignoring a large group of non-competitive overlaps would not serve the aim of distinguishing competitive and non-competitive speech on the basis of phonetic features alone; it would not benefit the search for discriminant use of prosody in discourse if one had first of all to identify backchannels on non-phonetic grounds, and remove them from the analysis. But still, our long-term aim is to progress towards a more refined automated analysis, starting by separating out instances of laughter, and possible other "emotional" displays. And in developing this research we will investigate further phonetic, and sequential, features than those reported here.

In summary, this paper contributes to ongoing efforts to apply prosodic research for applied, speech technological purposes, and highlights some of the challenges involved in this work. We question the extent to which prosody alone distinguishes dialogue functions, and conclude that the association between prosody and meaning is not as straightforward as is sometimes claimed, both with regard to conversational management, and more extensively to the display of emotions. One of the benefits of the current research is that we have been able to identify some of the exceptions to a commonly held claim. We have shown that there are other reasons for increasing loudness and pitch in overlap than those defined as turn-competition. Although this research did not succeed in demonstrating how particular features are used to distinguish dialogue acts, we maintain that there are great opportunities for applying at least some features of conversation analysis to large-scale analysis, and also, combining CA and automated techniques can lead to new methodological advances and discoveries.

Acknowledgments

References

Baayen, Rolf H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge, UK: Cambridge University Press.

Barth-Weingarten, Dagmar, Nicole Dehe & Anne Wichmann (eds.). 2009. *Where prosody meets pragmatics*. Bingley, UK: Emerald.

Breiman, Leo, Jerome Friedman, Charles J. Stone & Richard A. Olshen. 1984. *Classification and Regression Trees*. CRC press.

Cieslak, David A. & Nitesh V. Chawla. 2008. Learning decision trees for unbalanced data. Paper presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Antwerp, Belgium, 15-19 September.

Couper-Kuhlen, Elizabeth & Margaret Selting. 1996. *Prosody in Conversation*. Cambridge, UK: Cambridge University Press.

Deppermann, Arnulf. 2010. Future prospects of research on prosody: the need for publically available corpora: comments on Margaret Selting "Prosody in interaction: State of the art". In Dagmar Barth-Weingarten, Elizabeth Reber & Margaret Selting (eds.), *Prosody in Interaction*, 41-50. Amsterdam/Philadelphia: John Benjamins.

Edlund, Jens & Mattias Heldner. 2005. Exploring prosody in interaction control. *Phonetica* 62(2-4). 215–226.

Ford, Cecilia E. & Sandra A. Thompson. 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. *Studies in interactional sociolinguistics* 13. 134-184.

French, Peter & John Local. 1983. Turn-competetive incomings. *Journal of Pragmatics* 7. 17–38.

Glenn, Phillip. 2003. *Laughter in Interaction.* Cambridge, UK: Cambridge University Press.

Jefferson, Gail. 1984. Notes on some orderlinesses of overlap onset. In Valentina D'Urso and Paol Leonardi (eds.), *Discourse Analysis and Natural Rhetoric*, 11-38. Padua, Italy: Cleup Editore.

Kurtic, Emina. 2011. *Overlapping Talk and Turn Competition in Multi-Party Conversations*. Sheffield: University of Sheffield doctoral dissertation.

Kurtic, Emina, Guy J. Brown & Bill Wells. 2009. Fundamental frequency height as a resource for the management of overlap in talk-in-interaction. In Dagmar Barth- Weingarten, Nicole Dehe and Anne Wichmann (eds.), *Where Prosody Meets Pragmatics*, 183-204. Bingley, UK: Emerald Group Publishing Ltd.

Kurtic, Emina, Guy J. Brown & Bill Wells. 2010. Resources for turn competition in overlap in multi-party conversations: speech rate, pausing and duration. In *Eleventh Annual Conference of the International Speech Communication Association*, 2550–2553.

Kurtic, Emina, Guy J. Brown & Bill Wells. 2013. Resources for turn competition in overlapping talk. *Speech Communication* 55. 1-23.

Lerner, Gene H. 1996. On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker. In Eleanor Ochs &

Emanuel A. Schegloff (eds.), *Interaction and Grammar*, 238– 276. Cambridge, UK: Cambridge University Press.

Local, John. 2004. Getting back to prior talk: and-uh(m) as a back-connecting device in British and American English. In Elizabeth Couper-Kuhlen & Cecilia E. Ford (eds), *Sound patterns in interaction: cross-linguisitic studies of phonetics and prosody for conversation*, 377-400. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Local, John. 2005. On the interaction and phonetic design of collaborative completions. In William J. Hardcastle & Janet M. Beck (eds.), *A Figure of Speech: a Festschrift for John Laver*, 263-282. New Jersey, NY: Lawrence Erlbaum.

Oertel, Catharine, Marcin Wlodarczak, Alexey Tarasov, Nick Campbell & Petra Wagner. 2012. Context cues for classification of competitive and collaborative overlaps. In *Speech Prosody 2012*, 721-724. Shanghai, China.

Robinson, Jeffrey D. 2007. The role of numbers and statistics within conversation analysis. *Communication Methods and Measures* 1(1). 65–75.

Sacks, Harvey. 1992. *Lectures on conversation, volumes I and II*. In Gail Jefferson (ed.). Oxford, UK: Blackwell Publishing.

Schegloff, Emanuel A. 1982. Discourse as an interactional achievement: some uses of "uh huh" and other things that come between sentences. In D. Tannen (ed.), *Analyzing Discourse: Text and Talk* (71-93). Washington, DC: Georgetown University Press.

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

Selting, Margaret. 1996. Prosody as an activity-type distinctive cue in conversation: the case of so-called "astonished" questions in repair initiation. In Elizabeth Couper-Kuhlen and Margaret Selting (eds.), *Prosody in Conversation*, 231-270. Cambridge, UK: Cambridge University Press.

Selting, Margaret et al. 2011. Gesprächsanalyisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion, 10.* Available at: www.gespraechsforschung-ozs.de (accessed December 5, 2010).

Sidnell, Jack. 2011. *Conversation analysis: An introduction.* Vol. 45. John Wiley & Sons.

Szczepek Reed, Beatrice. 2009. Prosodic orientation: A practice for sequence organization in broadcast telephone openings. *Journal of Pragmatics* 41(6). 1223-1247.

Szczepek Reed, Beatrice. 2010. Prosody and alignment: a sequential perspective. *Cultural Studies of Science Education* 5(4). 859-867.

Szczepek Reed, Beatrice. 2011. Beyond the particular: prosody and the coordination of actions. *Language and Speech* 55(1). 13–34.

Szczepek Reed, Beatrice. 2015. Managing the Boundary Between "Yes" and "But": Two Ways of Disaffiliating With German ja aber and jaber. *Research on Language and Social Interaction* 48(1). 32-57.

Steensig, Jakob & Trine Heinemann. 2015. Opening up codings?. *Research on Language and Social Interaction* 48(1). 20-25.

Stivers, Tanya, Nicholas J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106(26). 10587-10592.

Stivers, Tanya & Nick J. Enfield. 2010. A coding scheme for question–response sequences in conversation. *Journal of Pragmatics* 42(10). 2620-2626.

Tracy, Karen. 2007. The role (or not) for numbers and statistics in qualitative research: an introduction. *Communication Methods and Measures* 1(1). 31–35.

Walker, Gareth. 2004. On some interactional and phonetic properties of increments to turns in talk-in-interaction. In Elizabeth Couper-Kuhlen & Cecilia. Ford (Eds.), *Sound Patterns in Interaction*, 147-169. Amsterdam, The Netherlands: John Benjamins.

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

Wichmann, Anne. 2011. Prosody and pragmatic effects. In Gisle Andersen & Karin Aijmer (eds.), *Pragmatics of Society* 5, 181-212. Göttingen, Germany: De Gruyter.

Zeitlyn, David. (forthcoming). Talking Kinship: doing kinship – some Mambila cases. (Kinship? Anthropology? Emergency! Emergency! Notes for a 21st Century Anthropology.). In Jon Abbink (ed.), *Rhetoric in Social Relations.* Oxford: Berghan Books.

Zuraidah, Mohd Don & Gerry Knowles. 2006. Prosody and turn-taking in Malay broadcast interviews. *Journal of Pragmatics* 38(4). 490-512.

Appendix

Gesprächsanalyisches Transkriptionssystem 2 (GAT 2); Selting et al. (2011)

<u>Sequence structure</u>
[ ]
[ ]   Overlaps
=   "Latching", a contribution starts immediately where a previous one ends

<u>Breathing</u>
°h / h°   In-breaths and out-breaths respectively, 0.2-0.5 sec
°hh / hh°   In-breaths and out-breaths respectively, 0.5-0.8 sec
°hhh / hhh°   In-breaths and out-breaths respectively, 0.8-1.0 sec

<u>Pauses</u>
(.)   Micro-pause, below 0.2 sec
(-)   Short pause, 0.2-0.5 sec
(--)   Medium pause, 0.5-0.8 sec
(---)   Longer pause, 0.8-1.0 sec
(1.0)   Longer pauses indicated by seconds

<u>Durations</u>
:   Prolongation of sound/syllable, 0.2-0.5 sec
::   Prolongation of sound/syllable, 0.5-0.8 sec
:::   Prolongation of sound/syllable, 0.8-1.0 sec

<u>Accents (in bold)</u>
ac´CENT   Rising pitch contour
ac`CENT   Falling pitch contour
ac¯CENT   Level pitch contour
acˇCENT   Falling-rising contour
acˆCENT   Rising-falling contour

<u>Pitch movement at the end of intonation phrases</u>
?   High rise
,   Middle rise
-               Level
;   Middle fall
.   Deep fall

<u>Other conventions</u>
ˀ   Glottal closure/hold
↑   Pitch step-up
↓   Pitch step-down
hu_uh   Two syllables
((head-move))   Non-verbal/non-spoken productions or events
(yes)   Candidate hearing
(he/you)   Possible candidates
<<p > >   Describing loudness and voice quality. Codes: p – piano, pp – pianissimo, f –
                forte, ff - fortissimo, all – fast, lento – slow