Resources for turn competition in overlap in multi-party conversations: Speech rate, pausing and duration

Emina Kurtić [1,2], *Guy. J. Brown* [2], *Bill Wells* [1]

[1] Department of Human Communication Sciences,
[2] Department of Computer Science
University of Sheffield, United Kingdom

emina@dcs.shef.ac.uk, g.brown@dcs.shef.ac.uk, bill.wells@sheffield.ac.uk

Abstract

This paper investigates the prosodic features that speakers use to compete for the turn when they talk simultaneously. Most previous research has focused on F0 and energy variation as resources for turn competition; here, we investigate the relevance of speech rate, pausing and the duration of in-overlap talk. These features are extracted from a set of overlaps drawn from the ICSI Meetings Corpus, and used to derive decision trees that classify overlapping talk as competitive or non-competitive. The decision trees show that both pausing and the duration of the in-overlap speech are significantly related to turn competition for both overlappers and overlappees. Additionally, speech rate is used by overlappees to return competition upon a turn competitive incoming. These findings partially support and extend the observations made in previous studies within the framework of conversation analysis and interactional phonetics.

Index Terms: overlapping speech, turn competition, prosody in talk-in-interaction, multi-party meetings

1. Introduction

Speakers normally exchange turns smoothly during conversation, talking one at a time. However, this practice sometimes breaks down, resulting in simultaneous talk and turn competition. Previous studies on turn taking and speaker overlap agree that overlapping talk can be – but doesn't have to be – an environment in which turn competition takes place. Accordingly, several prosodic features of overlapping speech which might be relevant to its characterization as turn-competitive or noncompetitive have been investigated.

These studies have identified fundamental frequency (F0) [1], energy [2] or a combination of the two [3] as resources that can be used by participants when competing for the turn. Furthermore, the relevance of overlap onset positioning [4] or its combination with the F0 and energy of competitive incomings [5] have been investigated, as well as speech rhythm around overlap onset [6] and hand motions [2].

Less consideration has been given to the role of speech rate, pausing and duration of in-overlap talk. Shriberg et al. [7] use speech rate features for an overlap-related classification task, although their study is not directly concerned with overlap competitiveness. They aimed to automatically classify words as in-overlap or in-the-clear based on their speech rate. They found that F0 and intensity were effective for this task, but speech rate was not. This suggests that there is little difference in the speech rate of in-overlap and clear talk, and hence that speech rate is unlikely to be a useful feature for turn competition.

However, different conclusions have been drawn from interactional studies. Schegloff [8] notes that competitive turn incomings are started in a fast tempo in order to circumvent competition, and that the responses to competitive incomings are characterized by slowing down the tempo and producing a sequence of stretched speech. French and Local [3] also note that decrease in rate can be used to return the competition. While turn holders continue at an unchanged pace upon a non-competitive incoming, a competitive incoming can be responded to by using slow tempo and increased loudness.

Regarding duration, Jefferson [4] notes that speakers sometimes keep on talking in overlap. In such cases, competitive overlaps are longer events in which both speakers persist beyond the point of realization that they are talking in overlap, and by doing so signal their interest in competing for the turn. Noncompetitive overlaps are expected to be shorter and resolved soon after such a point of realization.

Taking these previous observations as a starting point, this paper analyses speech rate, pausing and duration in overlap instances drawn from the ICSI Meeting Corpus [9]. We employ the method of conversation analysis (CA) to warrant the categorization of overlap as competitive or noncompetitive, based on observable orientation of conversation participants. Using features extracted from the ICSI recordings, decision trees are constructed that classify overlaps as competitive or noncompetitive. Our methodological assumption is that the use of single features and their combinations in competitiveness classification reflects how participants in conversation use these features for turn competition.

2. Methodology

2.1. Data

The analysis presented here uses a set of 1455 overlap instances (703 competitive, 752 noncompetitive) found in eight meetings[1] of the ICSI Meeting Corpus [9]. The meetings have between 6 and 8 participants, and they are all of a spontaneous type, in which there is an open exchange between participants. Given the multi-party nature of meetings, more than two participants can overlap at the same time. We only analyse overlaps in which two persons are involved and where both overlapper and overlappee are native speakers of American English. Furthermore, we remove from the analysis backchannels, collaborative and choral productions which have been shown to be mostly non-competitive [8]. We include only those non-competitive

[1] Bmr006, Bmr007, Bmr008, Bmr013, Bmr016, Bmr018, Bmr022 and Bmr025

overlaps which do not belong to one of these "default" non-competitive classes, and for which we can assume that it is their design in terms of features investigated here (and other prosodic features) that distinguishes them from competitive overlaps.

2.2. Overlap analysis

Sequences of overlapping talk were categorized as competitive or noncompetitive, using a CA approach (cf. [1]). For illustration consider the example given in Table 1.

A: we just- everything was just nice, so that- so the issue is in- in a situation
A: *.hhh* [where th- that's-]
B: *.hhh* [Well, it's not really] (.) nice.
B: It depends what you're doing.

Table 1: *Conversation extract from the ICSI meeting Bmr007 showing a turn-competitive overlap. The overlaps are indicated by square brackets ([]). '(.)' indicates a pause of duration less than 0.2s. and '.hhh' is the mark for inbreath [4].*

Speaker B starts her turn at the point in speaker A's turn that is not a point of syntactic completion, where it is clear that there is more to come. Even though B starts the overlap during the second part of A's turn (*so that – so the issue is...*), she chooses to address the first part of it (*everything was just nice*) thus attempting to stop A's turn. The analysis of a broader sequence of this conversation reveals that A's and B's *nice* do not refer to the same event, so the adjective is selected by B as a suitable reference point to attempt to interrupt A's turn and gain the floor for herself. A abandons his turn, whereupon B secures the floor for a lengthy turn of her own. These positional, syntactic and pragmatic criteria offer evidence of B's turn-competitive behavior, so this overlap is classified as competitive. Three annotators classified each overlap. The inter-annotator agreement, measured by Krippendorf's α, was $\alpha = 0.64$, comparable to that reported by [10] for a similar annotation task.

2.3. Feature computation

Speech rate (SR) is computed as number of syllables per unit of time. Syllables are determined from the phone-level forced-alignment available with the ICSI corpus and the syllable definitions in the CMU pronunciation lexicon, as used in the Festival speech synthesis system [11].

We express the SR of each speaker in two ways. Firstly, for both speakers we compute a speaker's in-overlap SR in relation to their normal SR, i.e. to the SR of in-the-clear turns. Secondly, we compute the difference in SR between the in-overlap talk and the talk immediately preceding the overlap. For overlappers, we compare the overlapper's in-overlap SR to the overlappee's SR prior to the overlap, to estimate whether matching or mismatching SR to that of the overlappee is a turn competitive device (as hypothesized by [6] for speech rhythm). For overlappees, we compare the overlappee's in-overlap SR to his SR just prior to the overlap onset, which gives us an indication of whether there is a change in SR upon overlap onset. We compute the overlappee's SR features over two units: the entire in-overlap context and the in-overlap context reduced by the "reaction time", a time span during which overlappees potentially realize that the overlap is underway, so that they tune their response accordingly. We set the reaction time to be the duration of the overlapper's first syllable, as previously suggested by [8]. This results in the following overlappee-related SR features:

- **eeSR_F1, eeSR_F2**: Overlappee's in-overlap SR normalized to the SR of that speaker's clear turns. eeSR_F1 is derived from the entire in-overlap context, and eeSR_F2 is derived from the in-overlap context reduced by the reaction time.
- **eeSR_F3, eeSR_F4**: Difference between overlappee's in-overlap SR and his pre-overlap SR. eeSR_F3 is derived from the entire in-overlap context and eeSR_F4 is derived from the in-overlap context reduced by the reaction time.

Pausing (PAU) features indicate the frequency, position and length of pauses in both speaker's in-overlap talk. In addition, for overlappers we include features that indicate whether the overlap is placed within or immediately after a pause, and give a length of such pause. It may be that pauses are interpreted as turn ends, and thus it is expected that non-competitive overlaps will be found in this position rather than competitive ones. However, this may depend on the pause length. The resulting feature set is as follows:

- **erPAU_F1**: Is overlap onset placed upon a pause in overlappee's turn?
- **erPAU_F2**: Length of the overlappee's pause upon which the overlapper starts the overlap
- **erPAU_F3**: Position of the first pause in overlapper's in-overlap talk
- **erPAU_F4**: Duration of the first pause in overlapper's in-overlap talk, normalized to the total pause duration in that speaker's clear segments
- **erPAU_F5**: Duration difference between overlapper's first in-overlap pause and overlappee's last pre-overlap pause
- **erPAU_F6**: Number of pauses in overlapper's in-overlap talk normalized to the number of words
- **erPAU_F7**: Total duration of all pauses in overlapper's in-overlap talk normalized to the total duration of in-overlap talk
- **eePAU_F1**: Number of pauses in overlappee's in-overlap talk normalized to the number of words
- **eePAU_F2**: Total duration of all overlappee's in-overlap pauses normalized to the total duration of the in-overlap talk
- **eePAU_F3**: Total duration of all overlappee's in-overlap pauses normalized to the total pause durations in that speaker's clear segments

Duration (DUR) related features express the duration of the in-overlap talk in terms of time and the number of words, capturing Jefferson's [4] observation that continuing to talk in overlap is a means to compete for the turn. For overlappers, an additional duration-related feature is included that describes the positioning of the overlap onset relative to the beginning of the overlappee's ongoing turn. This feature captures the expectation that competitive onsets are placed around the middle of the ongoing talk, while non-competitive ones will be positioned either towards the beginning or the end of overlapper's turn. This results in the following feature set:

- **erDUR_F1**: Duration between the beginning of overlappee's turn and the overlap onset, normalized by total duration of overlappee's turn containing overlap
- **erDUR_F2**: Duration (measured as number of overlapper's words) of the in-overlap section
- **erDUR_F3/ eeDUR_F2**: Duration (in time) of the in-overlap talk
- **eeDUR_F1**: Duration (measured as number of overlappee's words) of the in-overlap talk

2.4. Decision tree modelling

The role of the above features for turn competition was investigated using a classification paradigm. If the features are used in turn competition, then they will be good predictors of competitiveness when used to classify overlap instances from the ICSI data. Decision tree (DT) classifiers were used because the resulting trees allow a direct insight into the relevant features and their combinations.

To gain a full understanding of features' individual relevance for competitiveness classification, a *leave-one-in* approach was taken in which DTs were built using only one selected feature group. The performance of DTs trained on each single feature group was compared against a *majority baseline*

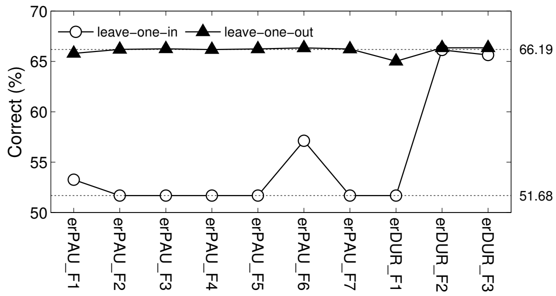

Figure 1: *Performance of single feature DTs for each feature from overlapper's PAU and DUR feature groups. Feature labels are explained in Section 2.3*

that classifies all overlap instances as non-competitive. The feature groups that lead to a performance significantly higher than the baseline were then further examined to show which specific features contributed to this result.

To analyze the relevance of feature combinations for competitiveness classification we use a *leave-one-out* method. We first build an all-feature DT, and then remove from it each feature group in turn. Changes in DT performance indicate the relevance of the removed feature group in combination with other features.

Classifier performance is assessed in terms of the percent of correctly classified overlap instances. All results in the next section are averaged over 10 repetitions of 10-fold cross-validation. Statistical significance is reported at the level $p < 0.05$ using the paired t-test.

3. Results

The first question we address is whether overlappers use the features discussed in the previous section to design their overlapping incomings as competitive, and thus differentiate them from non-competitive ones.

Feature group	*leave-one-in*	*leave-one-out*
ALL	66.19^*	
DUR	66.10^*	$57.88^\bullet$
PAU	56.87^*	65.64
SR	53.20	66.19

$*$: *significant change over the majority baseline (51.68%)*

$\bullet$: *significant change over the DT using all feature groups (66.19%)*

Table 2: *DT performance (% correct) for SR, PAU and DUR features describing overlapper's incoming.*

As shown in Table 2, DUR, PAU and SR features combined together (ALL) render a significant improvement in DT correctness over the majority baseline, which indicates that at least some of the features are good competitiveness predictors. DUR appears to be the most effective feature group. When used alone, duration significantly improves the classification performance over the majority baseline, and when omitted, the DT performance degrades significantly compared to ALL classifier. PAU features also significantly contribute to the classification when used alone, but can be omitted without loss of performance. SR features do not lead to a significant improvement over the majority baseline and are also redundant when used in combination with other features.

Figure 1 shows the performance of every single feature from the overlapper's DUR and PAU feature sets. Both DUR features lead to a significant improvement over the baseline, but are mutually redundant as their removal from the feature set results in no significant change in performance compared to the tree trained on all PAU and DUR features. The DT classifies all overlaps that are longer than 2 words or 0.53s as competitive, and the remaining ones as non-competitive. Among PAU features, the feature indicating the frequency of pauses (*erPAU_F6*) is the only feature that seems useful for competitiveness classification. Of all overlappers' incomings, 16.8% contain a pause, and if there is a pause, then the overlap is more likely to be competitive than non-competitive. Removing this or any other PAU feature from the total set of DUR and PAU features does not lead to a significant change in performance, which indicates that combining PAU and DUR features lends little to competitiveness classification. The tree that has access to all features makes competitiveness decisions based on overlap duration only.

Feature group	*leave-one-in*	*leave-one-out*
ALL	65.36^*	
DUR	65.95^*	$57.03^\bullet$
PAU	57.51^*	65.73
SR	57.41^*	65.24

$*$: *significant change over the majority baseline (51.68%)*

$\bullet$: *significant change over the DT using all feature groups (65.36%)*

Table 3: *DT performance (% correct) for SR, PAU and DUR features of overlappee's in-overlap talk.*

The next question we asked is whether overlappees respond to turn-competitive overlaps differently than to non-competitive ones in terms of overlap duration, pausing and speech rate. As shown in Table 3, the DT trained on all feature groups (ALL) achieves a significant improvement in correctness over the majority baseline, which is also the case for each of the three feature groups when used alone. The omission of durational features leads to a significant degradation in performance compared to the ALL tree, which indicates that durational features are the most dominant ones when used in combination with other feature groups.

Given that features from all three feature groups are potentially useful for competitiveness classification, we investigate the importance of each single feature. The result is shown in Figure 2. Comparable to the results for overlappers, pause frequency (*eePAU_F1*) and both DUR features (*eeDUR_F1* and *eeDUR_F2*) lead to significantly better results when used for classification alone. However, none of the SR features performs significantly above the majority baseline, although using all SR features together lead to a significant improvement. A closer inspection of the DT that uses only SR features reveals that a combination of two features that denote change in SR above the overlappee's norm (*eeSR_F1* and *eeSR_F2*) is useful for the classification, although neither of these features alone performs above the baseline.

Figure 2 shows redundancy between all individual overlappee's features that perform well in *leave-one-in* evaluation. Removing any of them does not significantly alter the performance of the entire feature group. The ALL DT makes use only of overlap duration, and classifies all overlaps longer than 0.6s or 2 words as competitive and non-competitive otherwise. A tree that has only PAU and SR features available uses the pause frequency and classifies all overlaps with 2 or

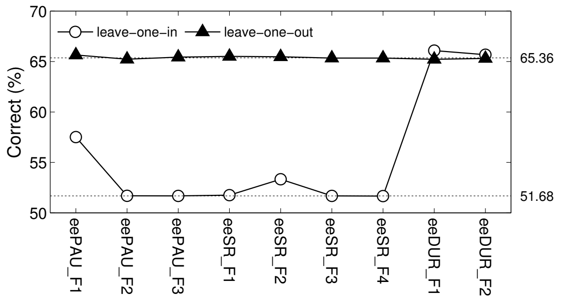

Figure 2: *Performance of single feature DTs for each feature from overlappee's PAU, DUR and SR feature groups. The feature labels are explained in Section 2.3*

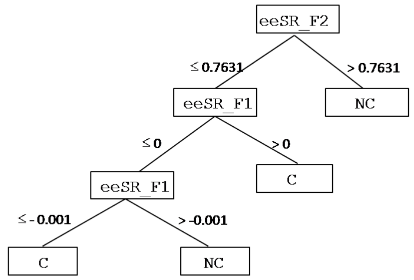

Figure 3: *Decision tree for overlappee's SR features. Labels are: competitive (C), noncompetitive (NC). The feature labels are explained in Section 2.3*

more pauses as competitive. The tree built from SR features only is shown in Figure 3. It shows that overlappees tend to increase SR above their norm for non-overlapping turns in reaction to a non-competitive overlap. In all overlaps where such an increase doesn't take place, the change in SR over the entire in-overlap section indicates that both competitive and non-competitive overlaps have an in-overlap SR that does not substantially differ from the norm. However, overlaps where overlappees are slower than the norm are more likely to be turn-competitive.

4. Conclusion

In this paper, speech rate, duration of in-overlap talk and pausing behaviour were investigated as resources that speakers use and orient to when competing for the turn in overlap. We observed that the most common resource that both overlapping speakers use for turn competition is simply continuing to speak in overlap beyond a threshold (as noted in [4]) which is approximately 2 words or 0.5s. Both speakers also pause more commonly in competitive overlaps, although the positioning and length of these pauses is not discriminative of competitiveness. Pausing can have multiple functions in competitive overlap, such as to monitor the progress of the competing talker. Deciding to continue upon a pause thus has a similar turn competitive function as continuing beyond a certain overlap duration. These two features are correlated, because longer overlaps are also more likely to contain pauses. This was apparent in our study, where pause features were found to be redundant if duration features were used by the DT.

Contrary to observations by [8], our study suggests that overlappers do not make use of faster speech rate to design the beginnings of their incomings as turn competitive. Nor does overlappers' modification of speech rate within the overlap discriminate between overlaps of different competitiveness. Overlappees may use speech rate to respond to turn competitive overlaps, although this feature is much less robust than duration and pausing. Overlappees tend to increase speech rate above their normal rate in noncompetitive overlaps and decrease it in competitive ones. The decrease upon realizing that a competitive overlap is in progress has previously been noted in [3] and [8]. An increase in speech rate as a response to a non-competitive incoming might reflect the desire for the turn-holder to resolve the overlap as soon as possible, and thus re-establish the *one-speaker-at-a-time* situation which is the default mode of turn taking.

In our future work we will use a similar paradigm to investigate the interaction between the features used here and F0, intensity and overlap onset position. We will also investigate the relationship between our findings here and the lexical and pragmatic properties of in-overlap turns.

5. References

[1] E. Kurtic, G. Brown, and B. Wells, "Fundamental frequency height as a resource for the management of overlap in talk-in-interaction," in *Where Prosody Meets Pragmatics (Studies in Pragmatics 8)*, D. Barth-Weingarten, N. Dehe, and A. Wichmann, Eds. Bingley:Emerald, 2009.

[2] C. Lee, S. Lee, and S. Narayanan, "An analysis of multimodal cues of interruption in dyadic spoken interactions," in *Proc. of Interspeech*, 2008, pp. 1678–1681.

[3] P. French and J. Local, "Turn-competitive incomings," *Journal of Pragmatics*, vol. 7, pp. 701–715, 1983.

[4] G. Jefferson, "A sketch of some orderly aspects of overlap in natural conversation," in *Conversation Analysis: Studies from the First Generation*, G. Lerner, Ed. Amsterdam: John Benjamins Publishing Company, 2003, pp. 43–59.

[5] B. Wells and S. Macfarlane, "Prosody as an interactional resource: Turn-projection and overlap," *Language and Speech*, vol. 41, no. 3–4, pp. 265–294, 1998.

[6] E. Couper-Kuhlen, *English Speech Rhytm: Form and Function in Everyday Verbal Interaction.* Amsterdam/Philadelphia: John Benjamins, 1993.

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

[8] E. Schegloff, "Overlapping talk and the organization of turn-taking for conversation," *Language in Society*, vol. 29, pp. 1–63, 2000.

[9] A. Janin, B. D., J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters, "The ICSI meeting corpus," *Proc. of ICASSP 2003*, 2003.

[10] M. Adda-Decker, C. Barras, G. Adda, P. Paroubek, P. de Mareüil, and B. Habert, "Annotation and analysis of overlapping speech in political interviews," *Proc. of LREC*, vol. 8, 2008.

[11] A. Black, P. Taylor, and R. Caley, *The Festival Speech Synthesis System*, 1st ed., 17th June 1999.