The Role of Speakers and Context in Classifying Competition in Overlapping Speech

Shammur Absar Chowdhury, Morena Danieli, Giuseppe Riccardi

Department of Information Engineering & Computer Science, University of Trento, Italy
{sachowdhury, danieli, riccardi}@disi.unitn.it

Abstract

Overlapping speech is one of the most frequently occurring events in the course of human-human conversations. Understanding the dynamics of overlapping speech is crucial for conversational analysis and for modeling human-machine dialog. Overlapping speech may signal the speaker's intention to grab the floor with a competitive *vs* non-competitive act. In this paper, we study the role of speakers, whether they initiate (*overlapper*) or not (*overlappee*) the overlap, and the context of the event. The speech overlap may be explained and predicted by the dialog context, the linguistic or acoustic descriptors. Our goal is to understand whether the competitiveness of the overlap is best predicted by the overlapper, the overlappee, the context or by their combinations. For each overlap and its context we have extracted acoustic, linguistic, and psycholinguistic features and combined decisions from the best classification models. The evaluation of the classifier has been carried out over call center human-human conversations. The results show that the complete knowledge of speakers' role and context highly contribute to the classification results when using acoustic and psycholinguistic features. Our findings also suggest that the lexical selections of the overlapper are good indicators of speaker's competitive or non-competitive intentions.

Index Terms: Spoken Conversation, Automatic Classification, Overlapping Speech, Discourse, Context

1. Introduction

Overlapping speech in spontaneous conversations is a naturally occurring phenomenon that may reveal speakers' attitudes, and in particular their intentions with respect to the control of the turn-taking structure of the conversations. In order to design conversational competent spoken dialog systems, the understanding of the overlapping phenomena is crucial. Over the years many linguists, psycholinguists, and speech researchers, have been studying these aspects of spoken interactions. In the conversational analysis tradition, overlaps have been considered as a violation of the fundamental rule [1, 2] of turn-taking, that is one person speaking at a time. Nevertheless, it has been shown that overlapping is pervasive in human conversations, for example authors in [3] suggest that about 40% of all between-speaker intervals to be overlaps. Further studies focused on highlighting speaker's intentions behind the overlaps. For example, it has been proposed that speech overlaps is related to dominance or aggression towards the other speaker [4]. However, the picture is more complex. Not all the overlapping occurrences are related to competitiveness. They also support cooperativeness in the conversations, for example in providing the other speaker with cues about the mutual understanding [5].

In the computational literature, a widely accepted categorization of overlaps, over years, is in between **Competitive (Cmp)**, *an attempt to grab the floor*, and **Non-Competitive (Ncm)**, *an attempt to assist the speaker for the continuation of the current turn*. Distinguishing the overlaps by the overlapper's intention is important for behavioral signal studies and for improving the quality of the spoken dialog system.

The aim of our study is to automatically classify competitive *vs* non-competitive overlaps. To classify overlaps, we focus on understanding whether the competitiveness is best represented by the information from speakers' segments: overlapper, overlappee, context or their combination. The roles of speakers and context conveying information varies with respect to different feature sets such as acoustic, linguistic, and psycholinguistic feature sets. For automatic classification, we investigated each speakers' segment enclosing overlaps and their combination for each feature set. We also combined the decisions from the best classification models, designed using different speakers' segments and feature sets, to obtain better classification results. For the experiments, we analyzed a large dataset of Italian spoken conversations collected in call centers, with customers and agents engaged in problem-solving tasks. Unlike most of previous studies, we investigate the role of speakers, context using linguistic and psycholinguistic features on 15,899 instances of overlaps.

The paper is organized as follows. An overview of previous studies of overlaps is given in Section 2, followed by the description and preparation of the dataset in Section 3. In Section 4, we discuss the details of the different speakers' segments, extracted features and classification experiments. Section 5 presents the results and analysis of our findings. Conclusions are provided in Section 6.

2. Related Studies

Few studies have been conducted on discriminating the speakers' competitive and non-competitive turns. Most of the studies indicates the importance of prosodic features, indicating fundamental-frequency (f0) and intensity as the dominant features [6, 7, 8]. In [9], the author suggests that the speakers raise their energy and voice when they attempt to interrupt the current speaker. In [10] a similar evidence is observed with pitch and amplitude.

Features such as speech rate, cut-offs and repetitions are also analyzed by conversational analysts. In [11], it is observed that to indicate competitiveness, variations in prosodic profiles and repetitions are used by speakers. The findings are also replicated for other languages in [12, 13] representing that in Italian human machine dialog repetitions and overlaps are not necessarily competitive but plays an important pragmatic role.

The study, in [14], showed that temporal features related to the position of overlaps and the onset-position of the overlap

are important for distinguishing the overlaps. Whereas in [6], the authors argued that the phonetic design plays an important role rather than its precise location for representing competitive overlap, which is also supported by the authors in [15, 10]. Duration of overlaps is found to be the most distinguishing feature while classifying competitive and non-competitive overlaps using decision tree [16, 17]. The authors also states that non-competitive overlaps tend to be shorter and resolved soon after the second speaker has recognized the overlap, and competitive overlaps are persistent because speakers keep on speaking despite overlapping.

For classifying overlaps different type of features has been explored, such as hand motion and disfluencies [18], body movement features from both speakers and contextual prosodic features from the overlapper [19], gaze, voice quality and contextual features –preceding and during overlaps [8]. The authors in [20] used different higher-dimensional acoustic feature types for categorizing overlaps, where they suggested that prosody and spectral features groups play an important role in characterizing competitive and non-competitive overlaps.

Aiming to predict interruptions, [21] found that interruptions are not random and context can be used to predict their occurrences. A similar conclusion is observed in [22], suggesting that interruptions are more likely to occur in intonational phrase units (IPUs) rather being random.

Unlike most of the previous studies, we aim to categorize competitiveness in overlaps by understanding the role played by different speakers' segments, enclosing an overlap, using acoustic, linguistic, and psycholinguistic features.

3. Data

3.1. Data Description

The data used for our research is a collection of Italian human-human spoken conversations, sampled from a large scale call center conversations providing customer care support. These conversations are recorded over two separate channels at a sample rate of 8 kHz, 16bits and have an average duration of 395 seconds. The corpus consists of 565 conversations with approximately 62 hours of data.

The manual segmentations and annotations of the speech overlaps with the competitive and non-competitive labels are performed by two Italian native expert annotators, using the guidelines described in [20]. The guideline includes competitive (Cmp) scenarios, where the intervening speaker (overlapper) starts prior to the completion of the current speaker (overlappee), both the speakers display interest in the turn for themselves, and the speakers perceive the overlap as problematic. As for Non-Competitive (Ncm) scenarios, the overlapper starts in the middle of an ongoing turn. No evidence is shown by both the speaker for grabbing the turn for themselves. The overlapper used the overlap to signal the support for the current speaker's continuation of speech. Both the speakers perceives the overlap as non-problematic.

For the annotations, the reported kappa measure between the annotators is 0.70. Examples of *Cmp* and *Ncm*, with their English translation, are shown in Table 1, where we also report tone direction, according to the notation in [23]. As shown in the example, in the *Ncm* scenario, the intention of S2 (the agent) is to repeat something that was already mentioned in the previous turns of the dialog: S2 wants to reassure S1 that she agrees on something that was already on the floor of the conversation; both overlapping segments are uttered with falling intonation.

Table 1: Dialog excerpts from the annotated corpus. Speech overlaps: bold form between [and], Hesitations: (.), Rising intonation: ↗, Falling intonation: ↘.

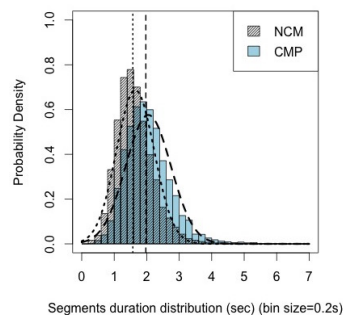| Ncm |
|---|
| S1: e quando [**cambiamo** ↘] (.) |
| S2:   [**sì sì** ↘ **ho già detto**] di cambiare ↘ |
| *S1: and when [**we change** ↘] (.)* |
| *S2:   [**yes yes** ↘ **I have already told**] to change ↘* |
| **Cmp** |
| S1: io non lo so [**io devo risparmiare**] ↘(.) |
| S2:   [**ma no la tariffa**] è buona ↗ |
| *S1: I do not know [**I had to save**] ↘(.)* |
| *S2:   [**but no the**] rate is good ↗* |

Figure 1: Duration distributions of overlap segments; NCM - non-competitive overlap segments; CMP - competitive overlap segments

On the contrary, in the *Cmp* scenario, S1 (the customer) is complaining about his problem, and he does not consider what S2 (the agent) claimed before. S2 has the intention to stop the complaints, and to take the turn from the on-going conversation, his overlapping segment has rising intonation and pitch level.

Table 2: Data set description.

Set	No. of Dialogs (% of Dialogs)	Duration	No. of Instances		Class Distribution	
			Cmp	Ncm	Cmp	Ncm
Train	341 (60.4%)	2 hrs 55 mins	2379	7158	24.94%	75.06%
Dev	109 (19.3%)	1 hrs 15 mins	724	2295	23.98%	76.02%
Test	115 (20.4%)	58 mins	763	2580	22.82%	77.18%

3.2. Data Preparation

For this study, we selected overlapping segments containing manual speech transcription. The exact boundary of the overlapping segments and their transcriptions are obtained using forced alignment between the word level transcriptions and the speech recording within the manual overlap segment boundary. For the alignment task, we used a domain specific automatic speech recognizer [24]. After forced alignment, we obtained 15,899 overlap segments, of a total duration of 5 hours and 8 minutes. The duration distribution of the overlapping segments is presented in Figure 1, where the median is shown using the dotted for Ncm and dashed for Cmp vertical lines. For the experiments, we split our data into train, dev and test sets. Details of the dataset are shown in Table 2.

4. Methodology

4.1. Overlapper, Overlappee, Context

In order to evaluate the roles of speakers and the context for classification of *Cmp vs Ncm*, we defined different speakers' segments enclosing overlaps, as shown in Figure 2, which are as follows:
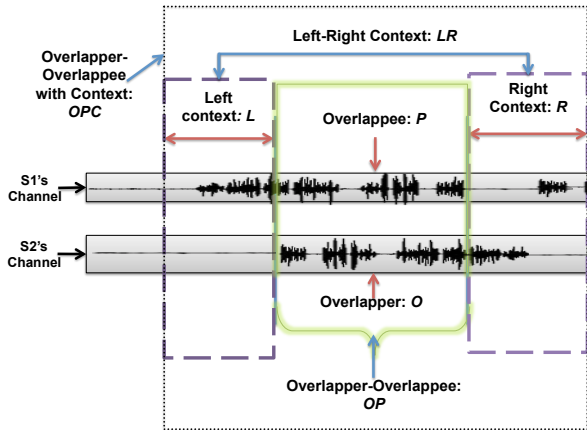
Figure 2: Example of different speakers' segments enclosing an overlap and their combination. S1-speaker 1; S2-speaker 2.

- Individual speakers' segments:
    - Overlapper (O): overlap initiator
    - Overlappee (P): current turn-holder
    - Left Context (L): speakers' segment before the start of the overlap
    - Right Context (R): speakers' segment after the completion of the overlap
- Combination of speakers' segments:
    - Overlapper-Overlappee (OP)
    - Left-Right Context (LR)
    - Overlapper-Overlappee with Left-Right Context (OPC)

One of the main challenges of studying about context is to decide the window size, which give us cues for classification/prediction. The author in [19], indicates that cues can be found in preceding segment (L) of overlapping speech but they do not exceed a window of 0.2s. The study also showed that window of 0.3s is sufficient for the following context (R).

From the manual annotation of context of our data we observed that the window size of the left context is $0.2s \pm 0.15s$ and right context is $0.8s \pm 0.5s$. We see that the window size of the right context varies a lot compared to the left context, which opens an avenue for further research. For this study, we used a window size of 0.2s and 0.3s, containing speech, for the left and the right context respectively, motivated by [19].

The left context (L) is defined by linearly merging speakers' channels where the information for speaker1's and speaker2's channels are $L_{s1} = \{l1_1, l1_2, ..., l1_m\}$ and $L_{s2} = \{l2_1, l2_2, ..., l2_m\}$ respectively. Then, we merged the features from both channels to form a new feature vector: $L = \{l1_1, l1_2, ..., l1_m, l2_1, l2_2, ..., l2_m\}$. The same procedure is used to design the feature vector for the right context (R).

The OP is designed by merging the overlapper (O) and overlappee (P) and the merged new feature vector is $OP = \{a_1, a_2, ..., a_m, b_1, b_2, ..., b_m\}$. We used similar approach to merge the left (L) and right (R) context to form LR.

In order to obtain the feature vector for Overlapper-Overlappee along with left and right context, we extracted features from both speakers' channels and merged them to obtain OPC, as same as OP. The boundary of the speaker channel is shown in Figure 2, which include overlap segments and contexts.

4.2. Features

Lexical features (Lex) We extracted lexical features using the boundary of start and end of the corresponding speakers' segments with forced aligned reference transcription. The lexical features are transformed into a bag-of-words (vector space model) [25]. The idea of the approach is to represent the words into numeric features. For this study, we extracted trigram features, to use the contextual benefit of n-grams, and selected the top 5000 frequent features to reduce the feature dimension.

Part-Of-Speech features (POS) We automatically annotated Part-Of-Speech tags using Tree Tagger [26]. After that, we used similar approach of lexical features for the transformation and reduction of the POS feature set.

Psycholinguistic features (LIWC) Psycholinguistic features are extracted from the transcription, using Linguistic Inquiry Word Count (LIWC) [27]. It has been used to study personality [28] among other social behaviors in order to understand the correlation between these attributes and word uses. The feature category include linguistic, psychological, personal concern (e.g., work, home), paralinguistic and punctuation among others.

Acoustic features (AC) We extracted low level acoustic features and then projected them onto statistical functionals using openSMILE [29], motivated by their successful utilization in several paralinguistic tasks discussed in [30]. The acoustic features are extracted with approximately 100 frames per second, with 25 milliseconds per frame. Acoustic features such as prosodic, spectral, voice quality, mfcc and energy are extracted. These low-level features along with its derivatives are then projected onto 24 statistical functionals like mean, range among others. More details of the features and the functionals are given in [20]. Unlike [20], we have not applied feature selection for different group of acoustic feature set in this study. In addition, the results of the acoustic feature sets are not comparable with [20], due to the different size of the datasets.

4.3. Classification and Evaluation

We trained our classification systems using Sequential Minimal Optimization (SMO), a support vector machine implementation of weka [31]. Prior to classification, feature values are normalized within $[0, 1]$ intervals. Due to the high-dimentionality of the feature vector and large number of instances, we used linear kernel of SMO with its default value of the penalty parameter, $C = 1.0$, for training the model.

For evaluation, we considered to measure Precision (**P**), Recall (**R**) and (**F1**), due to imbalanced class distribution. As we want to evaluate our system considering both of the classes, we computed macro-averaged P_{avg} and R_{avg}, which is an average of P and R for both classes, respectively. Using P_{avg} and R_{avg} we calculated the F1 for the overall system. For the simplicity, we are only reporting the F1 measure. Statistical significance has been reported in Section 5 using McNemar's test.

5. Results and Discussion

5.1. Classification Results

The performance of different speakers' segment and their associated feature set is reported in Table 3 for both dev and test set. For comparison, a SMO classifier has been designed using duration of overlapping segments as a feature for the baseline results. The baseline, **F1** for the dev and test set are 43.18 and 43.57, respectively. Results in Table 3 are significantly better compared to the baseline with $p < 0.001$.

We obtained best results with decision combinations as shown in the Table 3, with F1 69.41 and 66.43 on dev and test set respectively. The improved result in test set is significantly better compare to all of individual systems with $p < 0.001$. To combine the decisions from the best classification models we used majority voting ensemble method as a combiner. We selected four best models based on the performance of the dev set and the models are: 1) overlapper with lexical features ($O : Lex$), 2) overlapper-overlappee and context with acoustic features ($OPC : AC$), 3) overlapper-overlappee and context with psycholinguistic features ($OPC : LIWC$), and 4) overlapper-overlappee with POS features ($OP : POS$).

The system designed with lexical features from the overlapper channel performs better than any other individual system. The results on the dev set is 67.10 and on the test set is 64.99. The statistical significant test reveals that results using lexical features ($\boldsymbol{O : Lex}$) are highly significant with all other systems and their associated feature set ($p < 0.001$) except acoustic feature set in the context of $\boldsymbol{OPC}$. The $\boldsymbol{O : Lex}$ results are weakly significant compared to $\boldsymbol{OPC : AC}$ with a $p = 0.06$.

We are obtaining comparable results with acoustic features and it is an ideal condition when no transcriptions are available. The performance of the classifier designed with acoustic features extracted from overlapper-overlappee and context, $\boldsymbol{OPC}$, is F1 64.36 on the test set. In case of acoustic features, we observed that performance improves when we include context along with $\boldsymbol{OP}$.

As for the importance of context alone, the authors in [6] claim that no cues can be found before the overlaps. Our results with acoustic features from left context, $\boldsymbol{L}$, shows a similar characteristics. We obtained lower classification results, 52.44 of F1 on the test set. The lack of the contextual evidences affect on recall in case of *Cmp*. Another reason is the size of the left context window and in our case it is 0.2s of speech.

For the right context, the authors in [6] and [19] agree that the effect of competitive overlap sometimes get extended after the end of the overlap. A similar pattern is observed in our results using acoustic features of the right context, $\boldsymbol{R}$, where we obtained 9.89% improvement compared to the left context on the test set.

We observed that psycholinguistic features can distinguish competitive instances better when knowledge of the surrounding ($\boldsymbol{OPC}$) overlap is provided. One of the possible reasons is the presence of the change of word usage before, inside and following an overlap. We computed correlation coefficients between LIWC features and class labels using Pearson's correlation. We found that the highly correlated features are pronoun, cognitive processes, social processes among others.

The best performance with the POS features is observed in overlapper-overlappee and context (OPC) segment, giving a **F1** of 58.52 on the test set. However, with POS features extracted from overlapper-overlappee (OP) we obtained 61.12 on the dev set.

In summary, the competitiveness of the overlapping speech is best predicted using: 1) overlapper's lexical choice ($\boldsymbol{O}$:Lex), 2) acoustic and psycholinguistic features while exploiting the complete knowledge, i.e., speaker's role and context ($\boldsymbol{OPC}$:AC and $\boldsymbol{OPC}$:LIWC), 3) POS features when using overlapper information along with overlappee ($\boldsymbol{OP}$:POS), and 4) decision combinations of the best classification models.

5.2. Lexical Evidence

From the analysis of the token sequence uttered by the intervening speakers $\boldsymbol{O}$, we found that the lexical selections of the

Table 3: Classification Results for speakers' segments: overlapper O, overlappee P, left-context L, right-context R, along with the combination of overlapper-overlappe OP, left-right context LR and overlapper-overlappee with context OPC. Reported value is F1 measure of overall system. S.Seg: speaker's segment, Comb. Model: results for best model combination.

S.Seg	Eval	Lex	AC	POS	LIWC
O	Dev	*67.10*	62.80	59.73	53.60
	Test	**64.99**	60.38	59.37	52.60
P	Dev	59.50	58.43	55.73	50.10
	Test	58.87	57.01	55.35	50.05
L	Dev	50.77	52.05	49.37	50.00
	Test	50.47	52.44	51.62	50.05
R	Dev	52.04	62.34	51.24	50.05
	Test	51.27	62.33	49.35	49.85
OP	Dev	64.09	62.47	*61.12*	57.75
	Test	63.00	60.14	58.52	56.35
LR	Dev	52.07	59.82	48.26	50.00
	Test	51.50	59.40	48.31	49.75
OPC	Dev	64.27	*65.31*	59.02	*62.25*
	Test	62.57	**64.36**	**58.94**	**59.55**
Comb. Model	Dev	69.41			
	Test	**66.43**			

speakers may differ depending on their attitudes towards competitiveness. Comparing the frequencies of token for each class, we tested the statistical significance over the observed differences with a two-tailed two-sample t-test and $p = 0.1$.

Our findings suggests that in non-competitive instances, most frequent words indicates that the intervening speaker shares the opinions of the other speaker. For example, Italian words and phrases like "bene" ("well"), "ho capito" ("I have understood"), "certo" ("sure") are very frequent in *Ncm*. On the contrary, in the competitive distribution, occurrences of words and phrases like "no", "ma" ("but"), "mi scusi" ("excuse me") in Italian may play the role of discourse markers usually used to emphasize a discordant point of view. The findings implicates that for the non-competitive overlaps "sì" ("yes") is the most frequently used word to start an overlap, whereas the word "no", either alone or associated with adversative conjunctions like "ma" ("but"), is the most frequently used for competitive starts.

6. Conclusion and Future Work

In this study, we investigate the role of speakers and context using different speakers' segments, such as overlapper, overlappee, left and right context and their combinations. To understand their role, we employ linguistic, acoustic and psycholinguistic features. From the classification experiments we obtained best results while using decision combination method of the best models. The findings of our study suggest that the competitiveness of the overlapping speech is best classified using lexical choice of overlapper's turn; acoustic and psycholinguistic information from the overlapper-overlappe along with context. Future work will include a richer description of the context in terms of variable spans, semantic and discourse description.

7. Acknowledgments

8. References

[1] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language*, pp. 696–735, 1974.

[2] S. Duncan, "Some signals and rules for taking speaking turns in conversations." *Journal of personality and social psychology*, vol. 23, no. 2, p. 283, 1972.

[3] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversations," *Journal of Phonetics*, vol. 38, no. 4, pp. 555–568, 2010.

[4] C. West, "Against our will: Male interruptions of females in cross-sex conversation*," *Annals of the New York Academy of Sciences*, vol. 327, no. 1, pp. 81–96, 1979.

[5] J. A. Goldberg, "Interrupting the discourse on interruptions: An analysis in terms of relationally neutral, power-and rapport-oriented acts," *Journal of Pragmatics*, vol. 14, no. 6, pp. 883–903, 1990.

[6] P. French and J. Local, "Turn-competitive incomings," *Journal of Pragmatics*, vol. 7, no. 1, pp. 17–38, 1983.

[7] E. Kurtić, G. J. Brown, and B. Wells, "Resources for turn competition in overlapping talk," *Speech Communication*, vol. 55, no. 5, pp. 721–743, 2013.

[8] K. P. Truong, "Classification of cooperative and competitive overlaps in speech using cues from the context, overlapper, and overlappee," in *Proc. of INTERSPEECH*, 2013, pp. 1404–1408.

[9] E. Shriberg, A. Stolcke, and D. Baron, "Can prosody aid the automatic processing of multi-party meetings? evidence from predicting punctuation, dis uencies, and overlapping speech," in *ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding*, 2001.

[10] B. Hammarberg, B. Fritzell, J. Gaufin, J. Sundberg, and L. Wedin, "Perceptual and acoustic correlates of abnormal voice qualities," *Acta oto-laryngologica*, vol. 90, no. 1-6, pp. 441–451, 1980.

[11] E. A. Schegloff, "Overlapping talk and the organization of turn-taking for conversation," *Language in society*, vol. 29, no. 01, pp. 1–63, 2000.

[12] M. Danieli, C. Bazzanella, L. SpA, and I. Torino, "Linguistic markers in coming to understanding," in *Proceedings of VIII Meeting of AIIA (Associazione Italiana Intelligenza artificiale), AIIA 2002*, 2002, pp. 10–13.

[13] C. Bazzanella, *Repetition in dialogue.* Walter de Gruyter, 1996, vol. 11.

[14] G. Jefferson, *Two explorations of the organization of overlapping talk in conversation.* Tilburg University, Department of Language and Literature, 1982.

[15] 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.

[16] E. Kurtic, G. J. Brown, and B. Wells, "Resources for turn competition in overlap in multi-party conversations: speech rate, pausing and duration." in *Proc. of INTERSPEECH*, 2010, pp. 2550–2553.

[17] G. Jefferson, "A sketch of some orderly aspects of overlap in natural conversation," *PRAGMATICS AND BEYOND NEW SERIES*, vol. 125, pp. 43–62, 2004.

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

[19] C. Oertel, M. Wlodarczak, A. Tarasov, N. Campbell, and P. Wagner, "Context cues for classification of competitive and collaborative overlaps," *Proceedings of Speech Prosody 2012*, 2012.

[20] S. A. Chowdhury, M. Danieli, and G. Riccardi, "Annotating and categorizing competition in overlap speech," in *Proc. of ICASSP*. IEEE, 2015.

[21] C.-C. Lee and S. Narayanan, "Predicting interruptions in dyadic spoken interactions," in *Proc. of ICASSP*. IEEE, 2010, pp. 5250–5253.

[22] A. Gravano and J. Hirschberg, "A corpus-based study of interruptions in spoken dialogue." in *Proc. of INTERSPEECH*, 2012.

[23] I. P. Association, *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge University Press, 1999.

[24] S. A. Chowdhury, G. Riccardi, and F. Alam, "Unsupervised recognition and clustering of speech overlaps in spoken conversations," in *Proc. of Workshop on Speech, Language and Audio in Multimedia*, 2014.

[25] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Machine Learning: ECML-98*, ser. Lecture Notes in Computer Science, C. Nédellec and C. Rouveirol, Eds. Springer Berlin Heidelberg, 1998, vol. 1398, pp. 137–142. [Online]. Available: http://dx.doi.org/10.1007/BFb0026683

[26] H. Schmid, "Probabilistic part-of-speech tagging using decision trees," in *Proceedings of the international conference on new methods in language processing*, vol. 12. Citeseer, 1994, pp. 44–49.

[27] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, 2001.

[28] F. Alam and G. Riccardi, "Fusion of acoustic, linguistic and psycholinguistic features for speaker personality traits recognition," in *Proc. of ICASSP2014 - SLTC*, May 2014.

[29] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proc. of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 835–838.

[30] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Communication*, vol. 53, no. 9, pp. 1062–1087, 2011.

[31] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.