



# What can Speech and Language Tell us About the Working Alliance in Psychotherapy

S.P. Bayerl<sup>1</sup>, G. Roccabruna<sup>4</sup>, S.A. Chowdhury<sup>3</sup>, T. Ciulli<sup>2</sup>, M. Danieli<sup>4</sup>,  
K. Riedhammer<sup>1</sup>, G. Riccardi<sup>4</sup>

<sup>1</sup>Technische Hochschule Nürnberg Georg Simon Ohm, Germany

<sup>2</sup>Digital Psychology, IDEGO, Italy, <sup>3</sup>Qatar Computing Research Institute, Qatar,

<sup>4</sup>Signals and Interactive System Lab, DISI, University of Trento, Italy

giuseppe.riccardi@unitn.it

## Abstract

We are interested in the problem of conversational analysis and its application to the health domain. Cognitive Behavioral Therapy is a structured approach in psychotherapy, allowing the therapist to help the patient to identify and modify the maladaptive thoughts, behavior, or actions. This cooperative effort can be evaluated using the Working Alliance Inventory Observer-rated Shortened – a 12 items inventory covering task, goal, and relationship – which has a relevant influence on therapeutic outcomes. In this work, we investigate the relation between this alliance inventory and the spoken conversations (sessions) between the patient and the psychotherapist. We have delivered eight weeks of e-therapy, collected their audio and video call sessions, and manually transcribed them. The spoken conversations have been annotated and evaluated with WAI ratings by professional therapists. We have investigated speech and language features and their association with WAI items. The feature types include turn dynamics, lexical entrainment, and conversational descriptors extracted from the speech and language signals. Our findings provide strong evidence that a subset of these features are strong indicators of working alliance. To the best of our knowledge, this is the first and a novel study to exploit speech and language for characterising working alliance.

**Index Terms:** conversational analysis, working alliance, psychotherapy

## 1. Introduction

Research interest in automatic behavioral analysis has been increasing in the past few years [1]. It has been proved that carefully designed, AI-enabled, Personal Healthcare Agents (PHA) [2] may improve many aspects of the treatment processes, including citizens' access to healthcare services and improvement of patients' adherence to medical prescriptions and wellbeing recommendations.

To reach those benefits, it is crucial that the design and implementation of PHA may rely on the understanding of the human therapeutic process, collaborative design, and on the availability of high-quality resources like annotated corpora of patient-therapist verbal interactions. A critical factor for the success of the PHA is the patients' engagement during the care process. In other words, like in human-human patient-therapist interactions, the establishment of a satisfactory therapeutic relationship. The different facets of such relationships have been investigated by psychologists and medical doctors in depth in the recent past [3, 4]. The concept of therapeutic alliance has emerged as a candidate for operationally describing the multiple intertwined factors that characterize compliance with medical prescriptions, user satisfaction, and a high chance of reach-

ing benefits when setting and pursuing personal health goals. While this concept is general and applicable to all domains of healthcare, it is particularly crucial in mental healthcare because of the relevance of the bond that may be established between the therapists and their patients.

In this paper, we analyse patient-therapist spoken conversations collected over an eight-week Cognitive Behavioral Therapy (CBT) intervention in a real setting. The study aims at investigating the correlations between the conversation features and the patient-therapist working alliance. We describe the Working Alliance Inventory (WAI), the annotation process, and the inter-annotator agreement (IAA). The IAA results are very satisfactory and aligned with related studies using Cronbach's Alpha as IAA metric [5]. We have automatically extracted turn-taking, with other basic speech (e.g. duration and token), and language features from the conversations and correlated with WAI ratings. The turn-taking feature correlations show that equality in the engagement and acknowledgement utterances are relevant indexes of a good working alliance. The important role of the engagement is also observed in the most correlated turn-level features, which represent the speech activity at the segmentation and conversation levels. Furthermore, we have observed the same results on the discourse side, by measuring patient-therapist engagement through lexical entrainment for different word groupings and dialogue acts of these segments. Our analyses provides a strong evidence that the working alliance is manifested in speech and language. This result could provide support to further research in automatic behavioural analysis.

## 2. Concepts underlying therapeutic relationship

### 2.1. Working Alliance

The research on the evaluation of psychotherapy outcomes identified three key concepts that are frequently associated with good therapeutic results, measured in terms of satisfaction, acceptability, and symptom remission. The three concepts are proper information, cooperation, and commitment to achieve the objectives of treatment. While the concept of proper information pertains to the domain of ethics, cooperation and commitment to achieve treatment goals have been related to the construct of working (or therapeutic) alliance.

According to many authors, the working alliance are one of the strongest predictors of the successful outcome of a psychotherapeutic path [6, 7]. In [8], the study shows the association between Alliance and Adherence, defined as the extent

to which the patient’s behavior coincides with the medical or health advice.

The concept of working alliance has developed over time, and through the various approaches that characterize the psychological pathways. The authors in [9] explain how this concept has been discussed in the perspective of the different theories, from its origin in the psychodynamic approach, up to the contemporary view as a pantheoretical concept. Following a pantheoretical way, the study [10] identified the three underlying factors (related components). The three underlying factors of working alliance are the mutual understanding and trust that therapist and client may establish (bond), their willingness of setting a shared goal (goal), and the purposive collaboration between them (task).

Due to the importance of this construct in the determination of better results, over the years many researchers proposed questionnaires to measure the working alliance, for example, recently the authors in [5] studied the psychometric properties of this inventory for its application to self-assessment in online psychotherapy interventions.

## 2.2. Working Alliance Inventory

The Working Alliance Inventory (WAI henceforth, [11]) is a questionnaire of items arranged on a Likert scale and is one of the most used to assess the alliance [7] according to the integrative model by Bordin [10]. The three specific factors (bond, goal, and task) mentioned above were examined by a specific confirmatory factor analysis [12], which allowed defining the 12-item short form of the inventory. The questionnaire has been validated both for self-assessment and for being administered by an observer [13], and both versions have been validated in several languages. In this study, we administered the Italian form of the WAI Observer-rated Shortened (WAI-O-S) version [13] as reported in [14].

## 3. Corpus

### 3.1. Description

The corpus is a collection<sup>1</sup> of 23 Italian dyadic conversations during the sessions of Cognitive Behavioral Therapy (CBT). The intervention can last out from 5 to 20 sessions, each session of  $\approx 1$  hours, in which the patient works with the therapist on getting a better understanding of his/her issues. CBT therapy can reduce the impact on the patient’s life of many symptoms, such as phobia, depression, anxiety, and bipolar disorders. In this corpus, three patients attended, via voice and video call, 8 therapy sessions. However, for one of the patients, the first session is missing.

The patient-therapist dialogues were manually transcribed. In all, the corpus is composed of 5.9K turns, the dictionary size is 8.1K words and the average turn length is 12 tokens. The dialogues were segmented into turns using the silent pauses between utterances, thus multiple consecutive turns can belong to the same speaker.

### 3.2. Annotation task

The annotation task consisted of listening to the spoken conversation of a session and at the end of it, the annotators were asked to fill the WAI-O-S questionnaire by replying to the 12 questions listed in Table 1, by choosing a single option selection on

<sup>1</sup>This data collection has been approved by the Ethical Committee of the University of Trento.

Factors	Items
GOAL	<b>WAI 4:</b> There are doubts or a lack of understanding about what participants are trying to accomplish in therapy.
	<b>WAI 6:</b> The client and therapist are working on mutually agreed upon goals.
	<b>WAI 10:</b> The client and therapist have different ideas about what the client’s real problems are.
	<b>WAI 11:</b> The client and therapist have established a good understanding of the changes that would be good for the client.
TASK	<b>WAI 1:</b> There is agreement about the steps taken to help improve the client’s situation.
	<b>WAI 2:</b> There is agreement about the usefulness of the current activity in therapy (i.e., the client is seeing new ways to look at his/her problem).
	<b>WAI 8:</b> There is agreement on what is important for the client to work on.
	<b>WAI 12:</b> The client believes that the way they are working with his/her problem is correct.
BOND	<b>WAI 3:</b> There is mutual liking between the client and therapist.
	<b>WAI 5:</b> The client feels confident in the therapist’s ability to help the client.
	<b>WAI 7:</b> The client feels that the therapist appreciates him/her as a person.
	<b>WAI 9:</b> There is mutual trust between the client and therapist.

Table 1: WAI items and their categorization in factors [12].

a Likert scale from 1 (never) to 7 (ever). Two CBT experts performed the annotation of these therapeutic conversations. Based on those annotations, we computed the inter-annotator agreement (IAA) and the correlations to dialogue features.

The IAA has been computed with the Cronbach’s alpha [15] metric; according to its interpretation table [16], we report: 15 questionnaires achieve excellent internal consistency ( $\alpha > 0.9$ ), 4 good consistency ( $0.9 > \alpha \geq 0.8$ ), 1 acceptable ( $0.8 > \alpha \geq 0.7$ ), 1 questionable ( $0.7 > \alpha \geq 0.6$ ), 1 poor ( $0.6 > \alpha \geq 0.5$ ) and 1 unacceptable ( $\alpha \geq 0.5$ ). These values fall in line with other studies [5].

## 4. Conversational Features

Speech-based cues offer an important means of measuring cooperative efforts between the therapist and the patient. Moreover, these verbal cues hold the key to describe the turn-taking (interaction) dynamics between the interlocutors. Turn-taking is a remarkable phenomenon that represent the essence of human communication [17]. Over decades these intriguing cues are utilised to predict the outcome of conversations and dominance of speakers among others [18, 17, 19]. Motivated by its success to capture the flow of conversation, we exploit these features for understanding the working alliance between the therapist and the patient. For this study, we investigate the phenomena

in two stages using: (i) turn-taking measure incorporating conversational discourse and turn dynamics; (ii) turn-level measure where we studied windowed or individual segments in the conversation.

#### 4.1. Methods and Features

**Turn-taking Features** We extracted the turn-taking features by aligning outputs from speaker diarization sub-modules [20]. For the task, we utilised voice activity, overlap and speaker change detection using the raw dyadic audio. We aligned these automatic turn segments (short spurts/segments separated by at least 500ms silence) with the manual segmentation<sup>2</sup> and then assigned manually annotated speaker role.

Using these segments, we extracted global features to understand conversation structure and participation. The features includes:

1. **Participation equality** [21, 18] – measuring the difference in the proportion of turn duration hold by each speaker. We defined,  $P_{eq} = 1 - (\frac{\sum_i^N (T_i - T)^2}{E}) / T$  where  $T$  is the average speech duration of the speakers.  $T_i$  is the total speech duration for each speaker.  $E$  and  $N$  represent the maximum possible duration a speaker can hold and number of speakers. The value of  $P_{eq}$  closer to 1 indicate greater equality between speakers.
2. **Turn-level Freedom**, [21, 18] – measuring the predictability of turn-taking structure, defined by  $F_{cond} = 1 - \frac{H_{max}(Y|X) - H(Y|X)}{H_{max}(Y|X)}$ , where we calculated  $H(Y|X)$ , the conditional entropy of speaker  $Y$  being the next speaker after  $X$  begins the turn,  $H_{max}(Y|X)$  being the maximal possible value for this.  $W = \{therapist, patient\}$ ,  $X \in W$ ,  $Y \in W$  and  $X \neq Y$ . The value of  $F_{cond}$  is in between 0 and 1, where 0 represents a deterministic turn-taking (i.e. only speaker  $Y$  follows  $X$  strictly).
3. **Overlapping turns** [22, 19] – measuring the number and percentage of overlapping turns in the conversation.

**Turn-level Features** As the session unfolds, the turn properties of the speakers changes. We have observed large differences in positional features such as the mean turn duration of therapists and patients which ranged from 12.6 to 22.8 seconds for the therapist and 8.0 to 14.6 seconds for the patient.

Motivated by this, we extracted per speaker turn duration, token count, speech rates and pause-based (between and with-in the speakers) features for different parts (beginning/center/end), as well as the total percentage of speaking time for the interlocutors during the whole conversation like. For the study, we defined the beginning and end as the first and last 15% of the conversation (duration), the remaining part as center/middle position of the conversation.

#### 4.2. Results

The estimated Pearson correlation (with  $p$ -value  $< 0.05$ ), along with the regression slope, between the WAI-O-S ratings and the global turn-taking measures are reported in Table 2. The results indicate that an equal engagement ( $P_{eq}$ ) from both interlocutors can lead to better working alliance for Task (**WAI 1**, **WAI 2** and **WAI 12**), Goal (**WAI 11**) and Bond (**WAI 5**).

The result also pinpoints the more unpredictable turn-taking strategy the therapist used (for e.g., giving short feedback signals, showing agreement between patient's turns, or giving pa-

<sup>2</sup>Segmentation used for manual transcription.

Features	WAI (Factors)	Corr (Slope)
<i>Participation Equality</i>	WAI 1 (Task)	0.47 (2.59)
	WAI 2 (Task)	0.50 (4.16)
	WAI 5 (Bond)	0.42 (3.49)
	WAI 11 (Goal)	0.44 (3.36)
	WAI 12 (Task)	0.53 (4.72)
<i>Turn Freedom: Therapist*</i>	WAI 1 (Task)	0.43 (1.49)
	WAI 2 (Task)	0.54 (2.81)
	WAI 8 (Task)	0.53 (1.92)
	WAI 12 (Task)	0.51 (2.87)
<i>Minimal Speech Rate:</i>	WAI 9 (Bond)	0.69 (0.25)
<i>Therapist (tokens per sec)</i>	WAI 8 (Task)	0.60 (0.28)
<i>Number of Overlapping Turns</i>	WAI 2 (Task)	0.42 (0.03)
	WAI 5 (Bond)	0.42 (0.02)
	WAI 12 (Task)	0.47 (0.03)

Table 2: Reported significant Pearson correlation of turn-taking features ( $p$ -value  $< 0.05$ ) with WAI ratings and relative regression slopes (Slope).

tient more speaking time) often leads to better alliance scores for all the Task factors. Furthermore, we observed a slight positive effect of increase number of short overlapping spurt with Task (**WAI 2**, **WAI 12**) and Bond (**WAI 5**). This suggests that short feedback utterances are important for modelling a healthy alliance between the speakers.

Moreover, we observed a positive correlation between the minimum observed rate of speech of the therapist and the score for Task (**WAI 8**) and Bond (**WAI 9**). We hypothesize that the more mutual trust observed between patient and therapist, the therapist is more comfortable to talk faster. This also shows that the therapist is well-aware of the patient's agreement towards the treatment plan.

In addition to the global features, we also observed correlation with other positional turn information. We noticed, the median duration of the patient speaking in any observed time window is positively correlated ( $\rho = 0.64$ ) with the belief that the patient and therapist are working on his/her problems correctly (**WAI 12**). We hypothesize that the more regular the client engages in the conversation with the therapist, the more he/she is motivated to be working on his/her problems correctly.

We also observe a significant negative correlation ( $\rho = -0.55$ ) between the standard deviation of the therapists' turn duration and the rating for **WAI 11**. The more variation there is in therapists' turn duration, the less common understanding there is between patient and therapist on the changes that would benefit the patient. This could mean, long stretches of the therapist talking to the patient (schooling/telling) instead of talking with the patient, followed by a time that has more of a conversation characteristic. This hypothesis is also backed by the correlation coefficient seen between total percentage of turn duration belonging to patient ( $\rho = 0.58$  for **WAI 12** and  $\rho = 0.43$  for **WAI 11**).<sup>3</sup>

<sup>3</sup>A Negative correlation is seen with therapist's total percentage of turn duration.

There also is a positive correlation between the total duration the patient talks in the center part of the conversation (from 15% to 85% of the total duration, and the annotated rating for **WAI 5**. Meaning the more the client talks (engages in the conversation) during the middle part of the therapy session the more confidence he or she has that the therapist can help him or her.

## 5. Lexical Entrainment

The lexical entrainment is the phenomena in which a speaker conveys information reusing terms spoken by the other interlocutor. Furthermore, [23] defined a metric to estimate the entrainment, which effectiveness was tested on predicting the perceived naturalness of a dialogue.

The metric is defined as:

$$ENTR(c) = -\frac{\sum_{w \in c} |count_{S_1}(w) - count_{S_2}(w)|}{\sum_{w \in c} |count_{S_1}(w) + count_{S_2}(w)|} \quad (1)$$

where  $c$  is a target word class and  $count_{S_i}$  is the frequency of the word  $w$  used by the speakers  $S_1$  and  $S_2$ . The word frequencies are computed over the set of dialogues selected for the estimation of the entrainment. We have used the entrainment metric to analyze possible correlations with the normalized WAI ratings and entrainment rating. We hypothesize that WAI ratings, which assess the working alliance between patient and therapist, should have a relation with entrainment metrics measuring user engagement.

### 5.1. Methods and Results

In previous research [23], the word classes built to compute the entrainment were based on word frequencies, for instance, selecting the top 25 or 100 most frequent words of the whole corpus. In Table 3, we report the Pearson correlations between entrainment and WAI ratings for the top-25 most frequent words across the whole corpus (23 conversations). These words are reported in the first row of Table 3. We have further extended this analysis to all the words in the dictionary. We compute entrainment scores for each patient’s conversation. For each patient, we sort by frequency the words occurring in his/her conversations. We have then processed the sorted word list with a sliding window of 100 words and an overlap of 85 words. This process has generated 66 word-bins per patient. For each bin, we compute the entrainment scores (eq. 1) for each patient’s conversation. We correlate the entrainment scores, using Pearson correlation, with the 23 WAI ratings of a particular question. Due to the overall good inter-annotator agreement, we averaged the WAI ratings of the two annotators. Table 4 reports the highest correlation values amongst all word bins and WAI items. We have analyzed the part-of-speech of these words, using Spacy part-of-speech tagger, and have found 75% of the words are content words, i.e. they fall in one of these categories nouns, verbs, adjectives, adverbs, and main verbs. In the first column of Table 4, we show representative words in the bin for the corresponding WAI question. We have inspected these words and the context they appear. We manually categorize these words in terms of their discourse function according to the Dialogue Act standard ISO-24617-2 [24]. From this, we observe that *inform*, *agreement* and *offer* discourse functions correlate well with Task (**WAI 1**, **WAI 12**, **WAI 2**). While *feedback* and *request* discourse functions correlate with Goal (**WAI 11**) and Bond (**WAI 5**), respectively.

<b>Top-25 words</b>	“che”, “è”, “di”, “non”, “un”, “e”, “in”, “a”, “si”, “la”, “il”, “mi”, “per”, “anche”, “una”, “I”, “perché”, “quindi”, “sono”, “questo”, “se”, “ho”, “ci”, “le”, “ha”				
<b>WAI</b>	WAI 12	WAI 11	WAI 5	WAI 2	WAI 1
(Factors)	(Task)	(Goal)	(Bond)	(Task)	(Task)
<b>Corr</b>	0.52	0.46	0.46	0.45	0.44
(Slope)	(1.17)	(1.32)	(1.45)	(1.41)	(0.91)

Table 3: This table reports the top-5 Pearson correlations ( $p$ -value  $< 0.05$ ), along with regression slopes (Slope), between the entrainment ratings and WAI ratings. The word group used to compute the entrainment ratings includes the 25 most frequent words, listed in “Top-25 words” row, across the entire corpus.

Dialogue Act	WAI (Factors)	Corr (Slope)
<b>Inform:</b> “dicevi” (in therapeutic conversations it is used for promoting self-attention)	WAI 1 (Task)	0.64 (2.59)
<b>Agreement:</b> “infatti”, “effettivamente”, “assolutamente”	WAI 12 (Task)	0.64 (2.45)
<b>Offer:</b> “proviamo”, “vediamo”, “facciamo”, “possiamo”	WAI 2 (Task)	0.63 (3.41)
<b>Feedback functions:</b> “mah”, “beh”, “benissimo”, “esattamente”	WAI 11 (Goal)	0.62 (3.43)
<b>Request:</b> “pensa”, “aspetta”, “devi”	WAI 5 (Bond)	0.61 (3.71)

Table 4: Top-5 Pearson correlations ( $p$ -value  $< 0.05$ ), along with regression slopes (Slope), between WAI ratings and entrainments ratings of word groupings. The groupings were computed by a sliding window of 100 words wide with an overlap of 85 along the corpus word list sorted by frequency. In the column, “Dialogue Act” we select a sample of the most relevant words in that grouping and their associated dialogue act as per ISO-24617-2 standard [24].

## 6. Conclusions

We have presented the analysis of patient-psychotherapist conversations collected during an eight-week CBT intervention on three patients. We described the WAI assessment scheme of the working alliance, and then detailed the annotation task of these spoken conversations. Following, we analysed features extracted from speech and language cues to find correlations with the WAI ratings. We empirically showed that turn-taking structure in the session is evidently a predictor of WAI. We observed that encouraging participation equality and more patient turn time is a strong feature for higher WAI ratings. From the language side, we noticed higher use of *feedback*, along with *inform*, *request* discourse functions correlates significantly with higher WAI ratings. These results provide insights into the future work on automatically computing the working alliance in therapeutic sessions and extension to other health domains.

## 7. Acknowledgements

The research leading to these results has received funding from H2020 Grant Agreement 826266: COADAPT.

## 8. References

- [1] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1203–1233, 2013.
- [2] G. Riccardi, "Towards healthcare personal agents," in *Proceedings of the 2014 workshop on roadmapping the future of multi-modal interaction research including business opportunities and challenges*, 2014, pp. 53–56.
- [3] J. Pringle and K. C. Coley, "Improving medication adherence: a framework for community pharmacy-based interventions," *Integrated pharmacy research & practice*, vol. 4, p. 175, 2015.
- [4] S. Pihlaja, J.-H. Stenberg, K. Joutsenniemi, H. Mehik, V. Ritola, and G. Joffe, "Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders—a systematic review," *Internet Interventions*, vol. 11, p. 1, 2018.
- [5] R. Herrero, M. Vara, M. Miragall, C. Botella, A. García-Palacios, H. Ripet, A. Kleiboer, and R. M. Baños, "Working alliance inventory for online interventions-short form (wai-tech-sf): the role of the therapeutic alliance between patient and online program in therapeutic outcomes," *International journal of environmental research and public health*, vol. 17, no. 17, p. 6169, 2020.
- [6] A. O. Horvath and B. D. Symonds, "Relation between working alliance and outcome in psychotherapy: A meta-analysis." *Journal of counseling psychology*, vol. 38, no. 2, p. 139, 1991.
- [7] D. J. Martin, J. P. Garske, and M. K. Davis, "Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review." *Journal of consulting and clinical psychology*, vol. 68, no. 3, p. 438, 2000.
- [8] L. Thompson and R. McCabe, "The effect of clinician-patient alliance and communication on treatment adherence in mental health care: a systematic review," *BMC psychiatry*, vol. 12, no. 1, pp. 1–12, 2012.
- [9] A. O. Horvath and L. Luborsky, "The role of the therapeutic alliance in psychotherapy." *Journal of consulting and clinical psychology*, vol. 61, no. 4, p. 561, 1993.
- [10] E. S. Bordin, "The generalizability of the psychoanalytic concept of the working alliance." *Psychotherapy: Theory, research & practice*, vol. 16, no. 3, p. 252, 1979.
- [11] A. O. Horvath and L. S. Greenberg, "Development and validation of the working alliance inventory." *Journal of counseling psychology*, vol. 36, no. 2, p. 223, 1989.
- [12] T. J. Tracey and A. M. Kokotovic, "Factor structure of the working alliance inventory." *Psychological Assessment: A journal of consulting and clinical psychology*, vol. 1, no. 3, p. 207, 1989.
- [13] V. Tichenor and C. E. Hill, "A comparison of six measures of working alliance." *Psychotherapy: Theory, Research, Practice, Training*, vol. 26, no. 2, p. 195, 1989.
- [14] V. Lingiardi, *L'alleanza terapeutica: teoria, clinica, ricerca*. R. Cortina, 2002.
- [15] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.
- [16] J. A. Gliem and R. R. Gliem, "Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales." Midwest Research-to-Practice Conference in Adult, Continuing, and Community . . . , 2003.
- [17] Š. Beňuš, A. Gravano, and J. Hirschberg, "Pragmatic aspects of temporal accommodation in turn-taking," *Journal of Pragmatics*, vol. 43, no. 12, pp. 3001–3027, 2011.
- [18] S. A. Chowdhury, E. A. Stepanov, and G. Riccardi, "Predicting user satisfaction from turn-taking in spoken conversations," in *Interspeech 2016*, 2016, pp. 2910–2914.
- [19] S. A. Chowdhury, E. A. Stepanov, M. Danieli, and G. Riccardi, "Automatic classification of speech overlaps: feature representation and algorithms," *Computer Speech & Language*, vol. 55, pp. 145–167, 2019.
- [20] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, D. Fustes, H. Titeux, W. Bouaziz, and M.-P. Gill, "pyannotate.audio: neural building blocks for speaker diarization," in *ICASSP 2020, IEEE International Conference on Acoustics, Speech, and Signal Processing*, Barcelona, Spain, May 2020.
- [21] C. Lai, J. Carletta, and S. Renals, "Modelling participant affect in meetings with turn-taking features," in *Proc. Workshop of Affective Social Speech Signals*, 2013.
- [22] S. A. Chowdhury, M. Danieli, and G. Riccardi, "The role of speakers and context in classifying competition in overlapping speech," in *Sixteenth annual conference of the international speech communication association*, 2015.
- [23] J. B. Hirschberg, A. Nenkova, and A. Gravano, "High frequency word entrainment in spoken dialogue," 2008.
- [24] H. Bunt, J. Alexandersson, J.-W. Choe, A. C. Fang, K. Hasida, V. Petukhova, A. Popescu-Belis, and D. Traum, "Iso 24617-2: A semantically-based standard for dialogue annotation," in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 2012, pp. 430–437.