Coordination or Dominance? An Investigation of Social Dynamics in Conversational Entrainment

Nikolaos Flemotomos, Benjamin Ma, Raghuveer Peri
{flemotom,benjamjm,rperi}@usc.edu,
University of Southern California
Los Angeles, CA, USA

ABSTRACT

Speech entrainment is the tendency of individuals to adapt to their partners' characteristics over the course of a conversational interaction. While this is a well-studied phenomenon with direct social implications, it is usually defined in a symmetrical way, assuming that all the interlocutors involved in the interaction gradually become more similar to each other. In this project, we investigate whether entrainment is in fact manifested as an asymmetrical, directional relationship, where some participants dominate the conversation, while others try to adjust. We adopt a multimodal approach to measure entrainment and test the hypothesis that entrainment, when viewed through the lens of dominance and subordination, is affected by the group dynamics and the roles that the participants assume during the interaction.

ACM Reference Format:
Nikolaos Flemotomos, Benjamin Ma, Raghuveer Peri. 2021. Coordination or Dominance? An Investigation of Social Dynamics in Conversational Entrainment. In *Proceedings of CSCI 535: Multimodal Probabilistic Learning of Human Communication (CSCI 535).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Conversational entrainment is a well-studied phenomenon, according to which humans accommodate, or attune, their verbal and non-verbal communication style to their interlocutors. This can be observed through many features of communication, such as vocabulary, syntactic structure, speech rate, pitch, and gestures [8]. Entrainment is an important aspect of social interaction, since it reduces misunderstandings, strengthens social bonds, and, in general, enhances conversational efficiency [2].

Entrainment is often viewed as a convergence of communication behaviors, defined as "the situation where observed behaviors of two interactants, although dissimilar at the start of the interaction, are moving towards behavioral matching" [5]. This has been the basic theoretical foundation underlying many computational approaches to entrainment, modeling it as a linear phenomenon that can be quantified through similarity metrics based on correlation, recurrence analysis, and spectral methods [9]. It has been argued, though, that entrainment should be considered a dynamic process that can vary over the course of the conversation [2]. Tools have, thus, been proposed for incorporating the temporal dynamics to synchrony measurement [8]. Lately, neural embeddings able to capture entrainment characteristics at a local temporal scale have been introduced [19].

The vast majority of those methods are based on symmetric distance metrics, and are thus only able to answer the question of whether entrainment occurs, without explaining the potential directionality of the phenomenon. In other words, they do not take into account the fact that background factors such as personality and social status may influence the manifestation of entrainment. The Communication Accommodation Theory (CAT) [11] supports that individuals accommodate to their partners on a convergence-divergence continuum and suggests that the tendency to converge is proportional to one's need for social approval. Additionally, it was found in [3] that more dominant speakers tend to entrain less when compared with less dominant speakers. In related studies, it has been found that speakers' style converges more to interlocutors to whom they are positively disposed [24] or whom they perceive as belonging to a higher status [12].

The goal of this project is two-fold:

- Use multimodal (lexical and acoustic) information to analyze the effects of social status and formal roles (e.g., profession) on entrainment in multi-party conversations.
- Use directional distance metrics between conversational characteristics, in order to study not only the similarity, but also the dynamics between the speakers ("who is following whom").

The works most related to the proposed project are [7] and [4]. In particular, the authors in [7] study the relationship between entrainment and dominance in two different settings: in interactions between lawyers and justices at the Supreme Court (where justices have higher status and are thus expected to dominate the discussions), and in Wikipedia chat messages between page administrators and non-administrators. Beňuš et. al. extended those ideas providing additional evidence of directional entrainment at the Supreme Court [4]. However, only language-based analysis was considered in both, without accounting for acoustic-prosodic features for entrainment.

Additionally, very few existing works deal with multi-party dynamics (e.g., [23]). A study with respect to dominance and entrainment in multi-party interactions can be found in [16]; however, this study defines dominance as the variance of the participants' speaking times and therefore considers it a global undirectional group characteristic. Even though speaking time has a significant correlation with dominance, it has been shown that various other factors

CSCI 535, March 30, 2021, Los Angeles, CA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

affect perceived dominance dynamics within a small group [17]. We should here note that, after initial exploration of the task in hand, we decided to study multi-party entrainment as a set of entrainment scores between dyads.

2 DATASETS

Augmented Multi-party Interaction (AMI) meeting corpus [6] is a popular multimodal dataset consisting of several hours of audio and video recordings from multi-party meetings. In addition to far-field microphones, the dataset contains audio collected in close-talking scenario as well, including lapel-worn and headset microphones. In addition, high quality manual transcriptions are available for each individual participant. The corpus contains both scenario meetings, where each participant had a well-defined formal role (project manager, marketing expert, etc), as well as informal non-scenario meetings. In the scenario meetings, one out of the four participants was assigned the role of project manager (PM). The PM was designated to oversee the project from kick-off to completion. The role of the PM can be considered as a formal role, bestowed upon an individual owing to their designation. As mentioned previously, one goal of our work is to study the effect of such formal roles on entrainment. In addition, we also wanted to quantitatively analyze the effect of perceived dominance on entrainment in multi-party conversations. For this purpose, we have identified a subset of the AMI corpus described below:

DOME: Dominance in Meetings dataset[1] is a subset of the AMI corpus consisting of 11 sessions (roughly 4.5 hours of recordings). It consists of annotations of perceived dominance, particularly the most dominant (MD) and the least dominant (LD) participant, according to the annotators. The annotations were collected on non-overlapping five-minute meeting segments, resulting in a total of 58 segments, with each segment independently annotated by 3 annotators. This dataset would be useful to analyze the effect of dominance on entrainment. We computed the fleiss' kappa value to be 0.45, which can be considered moderate agreement [25]. However, considering only the MD and LD participants, majority agreement was found in 57 and 54 meeting segments respectively. Since the annotations provided are rankings of the participants from the least dominant to the most dominant on an ordinal scale, we use a modified version of Copeland's ranked voting scheme to compute scores for each participant. The Copeland score for a participant is the number of other participants over whom he or she has a majority preference[1]. We employ a modified version of this technique that preserves the distribution of ratings [18] to compute scores for each participant.

3 METHOD

3.1 Feature extraction

Expanding on the work of previous unimodal studies [4, 7], we study entrainment across two modalities: acoustic-prosodic and lexical.

3.1.1 Acoustic-prosodic. Previous studies have looked into several acoustic-prosodic features to measure entrainment [14, 15]. Drawing inspiration from these studies, we considered the intensity, pitch, jitter, shimmer, and noise-to-harmonics ratio (NHR) as features that capture vocal intensity, pitch, and quality. In particular, we follow the feature extraction method outlined in [15]: mean/max intensity, mean/max pitch, jitter, shimmer, and NHR values across 30 second, non-overlapping windows for each speaker in each meeting. A segment length of 30 seconds was chosen so that there is enough information to capture reliable features, while also providing enough number of samples for analyzing entrainment. In order to maintain high quality speech with minimal background noise, we used audio recordings collected using the lapel microphone, or in cases where it was not available we used the headset microphone. For analyzing entrainment on the DOME dataset, we extracted the features on the 5 minute segments for which the dominance labels were available. The feature extraction was performed using the Python speech processing library Parselmouth [13].

3.1.2 Lexical. In the lexical modality, we evaluate how each speaker's vocabulary usage changes during the conversation. Nenkova et al. [20] found that calculating lexical entrainment based on each speaker's frequency of using common words in the corpus led to significant predictive capability of task success and turn-taking. We adopt their approach to create lexical features in this study: we count the 25 most frequent words in the corpus, then compute each speaker's frequency of using each word as a separate feature. Word frequency $f(S_i, w)$ for speaker S_i on some word w is defined as:

$$f(S_i, w) = \frac{count_{S_i}(w)}{ALL_{S_i}} \tag{1}$$

where $count_{S_i}(w)$ is the amount of times S_i uses w in a given conversation span, and ALL_{S_i} is the total number of words spoken by S_i in that span. Before calculating word frequencies, we lemmatize all words using the WordNet Lemmatizer [10]. We also create two feature sets of word frequency features: one with stop words (e.g. "the," "it's") and filler words (e.g. "um," "mm-hmm") removed, and one with them included.

We additionally extract features using Linguistic Inquiry and Word Count (LIWC) categorizations [22]. Each word is classified into particular categories indicating word function or meaning (such as "affiliation", "personal pronoun", or "negative emotion") according to pre-defined LIWC category dictionaries. The way speakers choose and use specific lexical categories, such as function words (i.e., articles, prepositions, etc.) has been found to reflect important social characteristics including power and dominance dynamics [21]. We calculate LIWC features by counting occurrences of words used by each speaker in each of 92 LIWC categories. LIWC features were calculated at both turn-level ("local") and meeting-level ("global") scales (more details in the following section).

3.2 Measuring entrainment

The available role labels in AMI and its subset are given in both global (session-level) and local (short window-level) scales. In particular, formal roles (e.g., project manager vs. marketing expert) remain fixed throughout the entire meeting, while the dominance levels (i.e., most dominant vs. least dominant speaker) may change

[1] https://en.wikipedia.org/wiki/Copeland's_method

dynamically during the conversational interaction. Thus, we applied both global and local metrics of entrainment, proposed in the literature.

3.2.1 Global Metrics. A common way to measure entrainment globally is to study the absolute differences between specific acoustic or lexical features as estimated for different participants [15]. Formally, for a conversational trait T (e.g., mean pitch), a small value of $|T_{S_i} - T_{S_j}|$ indicates high level of entrainment between the speakers S_i and S_j ($i \neq j$). In order to confirm the expected tendency of speakers to entrain to each other, we study those differences in non-overlapping intervals corresponding to the beginning and ending of the conversation [16]. Those can be, for example, the first and second half of the recording. If a t-test reveals that the mean difference μ_T^I in the first interval differs significantly from the one in the second interval (μ_T^{II}), with $\mu_T^I > \mu_T^{II}$, it means that entrainment with respect to the trait T increases over the course of the conversation. In order to find a relationship between entrainment and dominant roles in that framework, we also estimate how much the same speech traits change for each speaker between the intervals. Our hypothesis, here, is that the dominant roles tend to maintain their style, while the non-dominant roles adapt.

3.2.2 Local Metrics. In order to study more subtle and dynamical changes of entrainment and how they relate to dominance, we utilized linguistic style coordination as defined in [7]. This is a more localized metric, studying entrainment at an utterance-by-utterance level employing words belonging in specific categories (e.g., articles, quantifiers, etc.) as lexical markers. In our implementation, those categories are extracted through the Linguistic Inquiry and Word Count (LIWC) dictionary [22]. Let's denote as u_S an utterance spoken by speaker S and as $u_{S,t}$ an utterance spoken by speaker S at some timestamp t, immediately after the utterance spoken at the timestamp $t-1$. Then, the linguistic coordination of B towards A with respect to some marker m is defined as

$$C_m(A \leftarrow B) = P(\mathcal{E}^m_{u_{B,t}} | \mathcal{E}^m_{u_{A,t-1}}) - P(\mathcal{E}^m_{u_B})$$

where $\mathcal{E}^m_u$ is the event that the utterance u contains a word from category m. Our hypothesis is that this directional metric of entrainment will be higher for less dominant speakers towards the more dominant ones.

4 EXPERIMENTS AND RESULTS

4.1 Global Audio Entrainment

In order to study audio-based entrainment, we first standardized all the extracted features per gender, and we estimated the average inter-speaker proximity during the first and last halves/quarters of each recording. Even though no statistically significant differences were observed using session halves as our intervals, statistical analysis through t-tests revealed that the proximity with respect to shimmer and jitter was significantly increased during the last quarter (at the $a = 0.05$ level after Bonferroni correction). As illustrated in Figure 1, the acoustic characteristics of PM remain relatively steady, and the other three participants accomodate their prosodic style to him/her. This validates our hypothesis that PM is a good proxy for the most dominant speaker and dominance can explain directional entrainment with respect to acoustic characteristics. We should note that we also tried to directly use the available dominance labels (DOME dataset), but we did not get any significant results. An explanation to that is the very small sample size (just 11 sessions available).

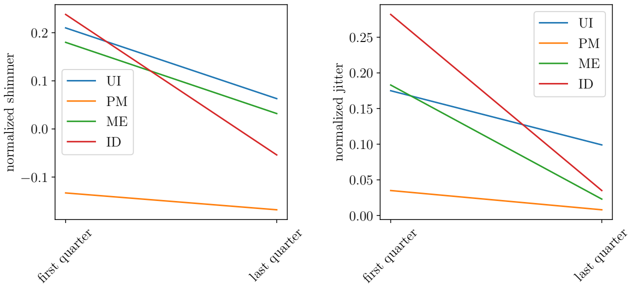

Figure 1: Average normalized shimmer (left) and jitter (right) during the first and last quarters of the available recorded meetings.

4.2 Global Lexical Entrainment

In order to study lexical entrainment at the global level, we used both the 25 most frequent words, and the LIWC categories as conversational traits. In the first case, we tried both keeping and removing stop words but in both scenarios the results were not conclusive. In particular, for some words we did not observe significant changes, while for some others we observed changes towards the opposite direction (less entrainment). We additionally constructed an aggregate entrainment metric, defined as the sum of individual word entrainment metrics, which was also observed to change towards less entrainment in the last quarter of the meeting.

However, when instead of studying individual words, we studied the LIWC lexical categories, and we estimated the average entrainment over all participants, we did observe significantly higher entrainment for 25 of the LIWC categories (at the $a = 0.05$ level after Bonferroni correction). Figure 2 illustrates how the linguistic patterns of the participants change over the course of the meeting with respect to the 5 categories with the most significant results. As shown, there are markers (e.g., conjunctions and prepositions) for which the Project Managers maintain their style while others converge towards them, but this is definitely not the case for all the categories.

4.3 Local Lexical Entrainment

To test the hypothesis that linguistic style coordination is higher for less dominant speakers towards the most dominant ones, after estimating the sample probabilities for all the markers for which enough data are available, we computed the mean coordination $\bar{C}(R_i \leftarrow R_j)$ across all the available markers for all the role pairs (R_i, R_j). Indeed, we found that $\bar{C}(\text{PM} \leftarrow R) > \bar{C}(R \leftarrow \text{PM})\ \forall R \in \{\text{ME,ID,UI}\}$. Additionally, we found that $\bar{C}(\text{ME} \leftarrow R) < \bar{C}(R \leftarrow \text{ME})\ \forall R \in \{\text{PM,ID,UI}\}$. Those results indicate that, according to the particular entrainment metric, the project manager (PM) is the most dominant role, while the marketing expert (ME) is the least dominant one. In Figure 3 we list the 5 LIWC categories with

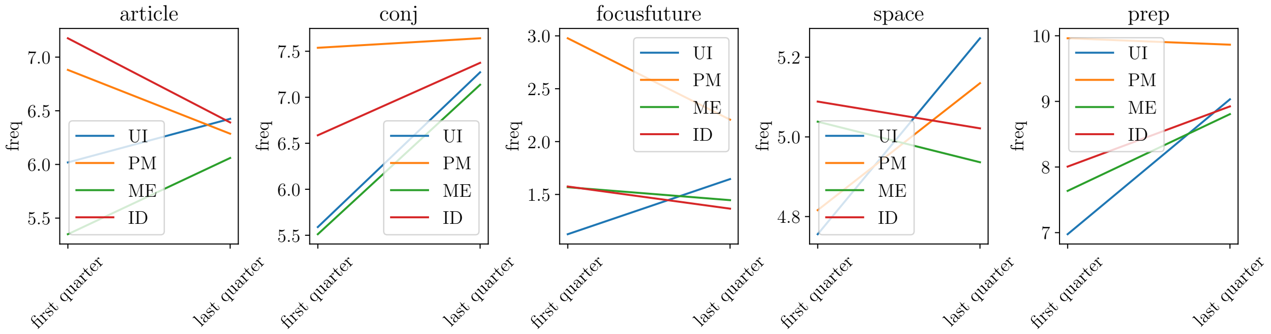

Figure 2: Frequency of words belonging to specific LIWC categories as used by the 4 participants during the first and last quarters of the available recorded meetings. The 5 categories with the most significant increase in average proximity are shown.

the largest differences between the coordination $C(\text{PM} \leftarrow R)$ and $C(R \leftarrow \text{PM})$ for all the role pairs. It is interesting to notice that, for all the pairs, categories like *social*, *affiliation*, *affect*, *pos*itive *emo*tion are consistently among the most significant ones. This tells us that when PM decides to make the meeting more cordial and informal, the rest of the participants accommodate their communication style to that.

5 CONCLUSION AND FUTURE WORK

In this project we studied the correlation between dominance and conversational entrainment and we established that the most dominant speakers maintain their linguistic and prosodic communication style, while the least dominant ones entrain towards them. Even though dominance labels are available for a subset of the dataset we are using, here we mainly focused on formal speaker roles as proxies for dominance. In the future we want to implement audio-based approaches to study entrainment at a local level and to correlate local entrainment with dominance. Additionally, we would like to use neural entrainment embeddings for the task in hand.

6 TEAM MEMBERS AND ROLES

The division of labor will be as follows:

(1) text-based feature extraction, feature normalization **Ben**
(2) audio-based feature extraction **Raghu**
(3) LIWC feature extraction, application of proposed entrainment metrics, statistical analyses **Nikos**
(4) final written report: **all**

REFERENCES

[1] Oya Aran, Hayley Hung, and Daniel Gatica-Perez. 2010. A multimodal corpus for studying dominance in small group conversations. *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality 18 May 2010* 22 (2010).
[2] Štefan Beňuš. 2014. Social aspects of entrainment in spoken interaction. *Cognitive Computation* 6, 4 (2014), 802–813.
[3] Š Benuš, A Gravano, and J Hirschberg. 2011. Pragmatic aspects of temporal entrainment in turn-taking. *J. Pragmat* 43, 12 (2011), 3001–3027.
[4] Štefan Beňuš, Agustín Gravano, Rivka Levitan, Sarah Ita Levitan, Laura Willson, and Julia Hirschberg. 2014. Entrainment, dominance and alliance in supreme court hearings. *Knowledge-Based Systems* 71 (2014), 3–14.
[5] Judee K Burgoon, Lesa A Stern, and Leesa Dillman. 1995. *Interpersonal adaptation: Dyadic interaction patterns.* Cambridge University Press.
[6] Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. 2005. The AMI meeting corpus: A pre-announcement. In *International workshop on machine learning for multimodal interaction.* Springer, 28–39.
[7] Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web.* 699–708.
[8] Céline De Looze, Stefan Scherer, Brian Vaughan, and Nick Campbell. 2014. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication* 58 (2014), 11–34.
[9] Emilie Delaherche, Mohamed Chetouani, Ammar Mahdhaoui, Catherine Saint-Georges, Sylvie Viaux, and David Cohen. 2012. Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing* 3, 3 (2012), 349–365.
[10] Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database.* Bradford Books.
[11] Howard Giles, Nikolas Coupland, and Justine Coupland. 1991. *Accommodation theory: Communication, context, and consequence.* Cambridge University Press, 1–68. https://doi.org/10.1017/CBO9780511663673.001
[12] Stanford W Gregory Jr and Stephen Webster. 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of personality and social psychology* 70, 6 (1996), 1231.
[13] Yannick Jadoul, Bill Thompson, and Bart de Boer. 2018. Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics* 71 (2018), 1–15. https://doi.org/10.1016/j.wocn.2018.07.001
[14] Rivka Levitan, Agustin Gravano, and Julia Hirschberg. 2011. Entrainment in Speech Preceding Backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.* 113–117.
[15] Rivka Levitan and Julia Hirschberg. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Twelfth Annual Conference of the International Speech Communication Association.*
[16] Diane Litman, Susannah Paletz, Zahra Rahimi, Stefani Allegretti, and Caitlin Rice. 2016. The teams corpus and entrainment in multi-party spoken dialogues. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing.* 1421–1431.
[17] Marianne Schmid Mast. 2002. Dominance as expressed and inferred through speaking time: A meta-analysis. *Human Communication Research* 28, 3 (2002), 420–450.
[18] Karel Mundnich, Md Nasir, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2017. Exploiting Intra-Annotator Rating Consistency Through Copeland's Method for Estimation of Ground Truth Labels in Couples' Therapy.. In *INTERSPEECH.* 3167–3171.
[19] Md Nasir, Brian Baucom, Craig Bryan, Shrikanth Narayanan, and Panayiotis Georgiou. 2020. Modeling vocal entrainment in conversational speech using deep unsupervised learning. *IEEE Transactions on Affective Computing* (2020).
[20] Ani Nenkova, Agustín Gravano, and Julia Hirschberg. 2008. High Frequency Word Entrainment in Spoken Dialogue. *Proceedings of the ACL/HLT 2008*, 169–172. https://doi.org/10.3115/1557690.1557737
[21] Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology* 21, 4 (2002), 337–360.

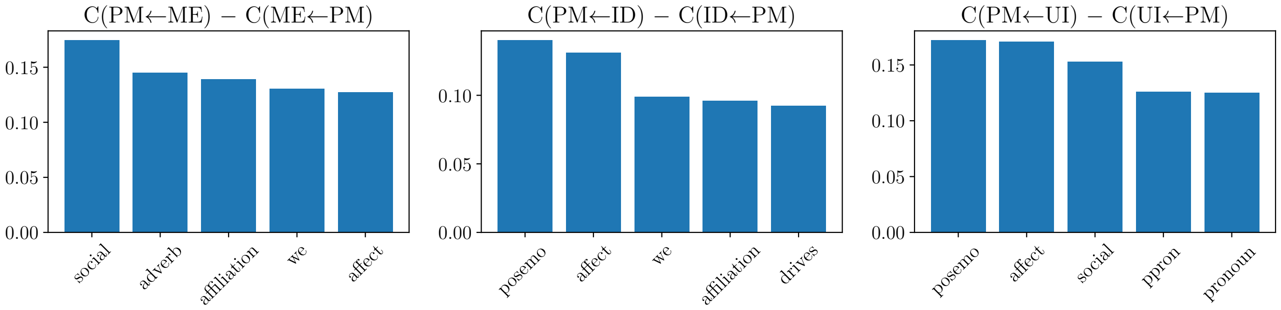

Figure 3: The 5 LIWC categories/markers yielding the largest difference between the linguistic coordination $C(\textbf{PM} \leftarrow R)$ and $C(R \leftarrow \textbf{PM})$ for all the other available roles R.

[22] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015.* Technical Report.

[23] Zahra Rahimi and Diane Litman. 2020. Entrainment2Vec: Embedding Entrainment for Multi-Party Dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 8681–8688.

[24] Alan Yu, Carissa Abrego-Collier, Rebekah Baglini, Tommy Grano, Martina Martinovic, Charles Otte III, Julia Thomas, and Jasmin Urban. 2011. Speaker attitude and sexual orientation affect phonetic imitation. *University of Pennsylvania Working Papers in Linguistics* 17, 1 (2011), 26.

[25] Massimo Zancanaro, Bruno Lepri, and Fabio Pianesi. 2006. Automatic detection of group functional roles in face to face interactions. In *Proceedings of the 8th international conference on Multimodal interfaces.* 28–34.