“Was that your mother on the phone?”: Classifying Interpersonal Relationships between Dialog Participants with Lexical and Acoustic Properties

Denys Katerenchuk[1], *David Guy Brizan*[1], *Andrew Rosenberg*[1,2]

[1]CUNY Graduate Center, New York, USA
[2]CUNY Queens College, New York, USA
{dkaterenchuk, dbrizan}@gc.cuny.edu, andrew@cs.qc.cuny.edu

10.21437/Interspeech.2014-416

Abstract

Understanding interpersonal relationships provides important context in understanding spoken communication. In addition to increasing knowledge of the social indicators in spoken communication, the automatic recognition of interpersonal relationships has an application in providing structure to social networks. This paper presents exploratory work on the challenging problem of distinguishing family from friends in spontaneous dialogs drawn from the CALLHOME English corpus. We find both acoustic/prosodic and lexical features useful in classifying these relationships. In binary classification experiments, we achieve accuracy of 10.71% absolute improvement over chance (50%) assignment.

Index Terms: speech understanding, interpersonal relationships, information extraction

1. Introduction

You are in a meeting when the phone rings. Your boss excuses herself, picks up the phone, has a brief conversation, apologizes for the interruption, and your meeting continues. Can you tell if your boss is talking to her mother? brother? partner? colleague?

You are on the subway. You hear two men talking. You listen to their conversation for a few minutes. They catch up and exchange stories about people they both know for a few minutes until you arrive at your stop and get off the train. Were they brothers? Father and son? Strangers who just met?

People are remarkably capable of assessing the relationship between two speakers, and even the relationship between a single speaker and his or her unheard interlocutor. In this paper, we explore this question by analyzing conversations in the CALLHOME English [1] corpus and determining the ability to automatically classify interpersonal relationships based on lexical and acoustic properties of the conversation. Specifically, we are examining the question of if we can distinguish conversations between friends from conversations between family members.

This is a challenging problem. While humans are remarkably good of accomplishing this task, the distinction between how two close friends communicate and how, say, siblings or cousins communicate can be subtle. Despite this similarity Patrick and Metcalf found that communication in family and friends can manifest itself differently, with familial communication being more goal-oriented and obligatory, while friendly communication is more often low-intensity “chit-chat” [2].

This work is related to previous work (cf. Section 2) which aims to increase understanding the social functions of language. In contrast to work that examines social roles and power dynamics, here we are investigating the ways which familial and friendly relationships differ in their spoken communication. In addition to increasing our understanding of human communication style and strategy, there are applications of this kind of automatic classification in understanding dialog speech in observed conversations. Understanding the interpersonal relationships between interlocutors can provide important context for interpretation. In the context of social network analysis, interpersonal relationship information can be used to annotate the links between people.

In this paper, we describe the CALLHOME English corpus and our annotation of interpersonal relationships (cf. Section 3). We then present the acoustic and transcript-derived lexical features used to classify these relationships (cf. Section 4). In Sections 5.1 and 5.2, we present and discuss classification results based on full conversations and contrast this with classification based on a single side of a conversation. We conclude and provide directions for future work in Section 6.

2. Related Work

There has been a substantial amount of interest in analyzing social information from spoken and text based interpersonal communication using corpus analysis and automatic classification. Much of this is in the context of analyzing speaker roles. On broadcast news (BN) and broadcast conversation (BC) speech, this is typically cast as distinguishing roles such as anchors, reporters, and guests [3, 4, 5]. Additional related work is in the context of meeting data on both the ICSI [6] and AMI Meeting [7] corpora. In the AMI Meeting corpus, participants took on the roles of project manager, marketing expert, UI designer and industrial designer. In the ICSI Meeting data, meeting participants are either Advisors, Students, Staff or Guests. Research has been done on both of these corpora analyzing lexical cues, social network cues, acoustic information, and turn-taking behavior [8].

In contrast to these, Sapru and Valenti makes a distinction between these formal roles and the social roles, protagonist, supporter, gatekeeper, neutral, and attacker [9]. This work found that lexical features were most useful for distinguishing formal roles, while prosodic information was most effective in distinguishing social roles.

Campbell [10] examined the speech of a single Japanese speaker as she spoke to different conversational partners, finding significant differences to pitch (F0) and normalized amplitude quotient (NAQ) depending on the relationship between the speaker and her conversational partner.

In a similar vein, other work has investigated power rela-

14 – 18 September 2014, Singapore

tions and intimacy. Fitzsimmons and Kay found that the use of the pronoun "we" to be important in establishing closeness [11]. Danescu-Niculescu-Mizli et al. investigated the use of language as relating to power relationships in Wikipedia communities and US Supreme Court arguments. In both they find that power is reflected in how quickly a person adopts the linguistic style of the person to whom they are responding [12]. Similarly, Ireland et al. found that similarity in linguistic style predicted mutual attraction in speed dating [13].

Also in the speed-dating domain, Ranganath et al. examined acoustic/prosodic and lexical correlates of interpersonal stances like awkwardness and flirtatiousness [14]. In addition to the frequency of pronouns and other common words, they examined the impact of discourse and dialog features, including interruptions, restarts, agreement and sympathy. This analysis also included rates of accommodation – the degree by which interlocutors' use of language resemble each other.

While it is clearly a wealth of work investigating the social aspects of language and communication, we believe this preliminary study is the first work to automatically distinguish friends from family members on the basis of their communication.

3. Data

As described in Section 2, the majority of work in the classification of interpersonal relationships has been in the context of formal social roles in broadcast news and meetings. In contrast, we want to predict relationships from spontaneous conversations. This lack of available annotated data for this task led us to investigate the CALLHOME English corpus.

This corpus is a collection of native English speech telephone conversations including 120 conversations, each up to 30 minutes long, for a total of 56.7 hours of spontaneous dialog. The conversations are divided into train (80 conversations), development test (20) and test (20) sets. The speakers are representatives of various demographic groups. Most conversations were between friends and relatives. The subset of the corpus has been manually transcribed. This includes 10 minutes of transcribed speech for the train and development test sets, and five minutes of transcribed speech for the evaluation set totaling 18.3 hours of transcripts.

The annotation of interpersonal relationships between conversational partners is not part of the CALLHOME English corpus. These annotations were performed by a group of seven annotators from the Speech Lab @ Queens College. Annotators could listen to the full conversation and refer to the transcripts where available. Annotators could describe the relationship using any term they liked. However, all annotations were entered into a shared document; this led to a relatively rapid convergence to a small set of labels. That said, there was still some individual differences in the labels that were resolved after annotation. We found that most conversations were between friends – some of whom could be identified as work colleagues. However, we found that a finer grained distinction of *types* of friends could not be reliably determined across the whole corpus.

We ultimately settled on two binary interpersonal relationship labels, FRIENDS and FAMILY. The line between these groups is very thin since very close friends feel like relatives; also, cousins and siblings can be friends. Labeling decisions were based on both audio and transcripts. We found this task to be non-trivial in many cases. We were unable to provide labels for 12 conversations (10% of the corpus) because 1) the relationship could not be determined with confidence or, 2) in two instances, more than two speakers joined the conversation. An interesting quality of the CALLHOME data is that a small number of the conversations were between nuns who refer to each other as "sisters," when they were actually friends or colleagues.

In the labeled CALLHOME corpus, the majority of instances, 80 out of 108, are FRIENDS. This creates highly unbalanced corpus. In our experiments we use a subset of this corpus that was trimmed down to 56 instances comprising all 28 FAMILY conversations (15 between siblings, 13 between parents and children) and 28 randomly chosen FRIENDS conversations. These annotations of the CALLHOME English corpus are available at `http://speech.cs.qc.cuny.edu`.

4. Methods

In this section, we describe the acoustic and lexical features we use to classify interpersonal relationships from conversations.

Using the class-balanced subset of conversations, we perform all classification experiments using ten-fold cross-validation. No speaker occurs in more than one conversation, and no conversation appears simultaneously in a train and test partition.

We derive a number of acoustic and transcript-based features in order to predict the relationship of the interlocutors using Weka [15]. The features are described below.

4.1. Acoustic Features

People communicate differently depending on the relationships. The way a person talks to the parents is different from the way he or she talks to a friend. Not only the word choice, but also intonation and stress will vary. This is the reason we look into acoustic/prosodic features.

Our acoustic feature extraction routine is based on the INTERSPEECH (IS) 2009 Emotion Challenge [16]. We use open-source feature extraction openSMILE [17]. The configuration file is a version of the emotion challenge, IS09_emotion.conf, that was distributed at IS 2009 and publicly distributed with openSMILE, modified to output arff files with appropriate labels. The modified configuration file is available at `http://speech.cs.qc.cuny.edu/`; the original is distributed via `http://sourceforge.net/projects/opensmile/files/`. We extract 384 features from the audio. This feature set has been shown to perform well on initial test experiments compared to other publicly-distributed configuration recipes. In particular, the IS 2010 paralinguistic and emotion challenge feature sets, with 1,582 and 6,552 features respectively, showed lower performance.

OpenSMILE is used to extract acoustic features from the full conversation. The openSMILE feature extraction process operates by first extracting a set of short frame-based low-level descriptors (LLDs) and then applying functionals over these descriptors to extract aggregated features.

This feature set includes five LLDs: 1) Zero crossing rate, 2) RMS Energy, 3) F0, 4) Harmonic-to-Noise Ratio, and 5-16) 12 MFCC coefficients. The change (Δ) of each of these LLDs is also calculated. This leads to a total of $16 \cdot 2 = 32$ LLDs. Twelve functionals are then applied to these: 1) mean, 2) standard deviation, 3) skewness, 4) kurtosis, 5-8) value and relative position of minima and maxima, 9) range between minima and maxima, 10-12) linear regression coefficient, offset and MSE.

4.2. Transcript-derived Features

Based on the transcriptions of the conversations [1], we extracted a number of text-based features. Chung and Pennebaker

[18] suggest that the analysis of function words (pronouns, articles and other closed-class words) can reveal the emotional state of a person, whether the person is stilted, hedging or enthusiastic, for example. Function words also do not suffer from sparse data problems; they account for so many of the tokens in language that they are present in almost all contexts.

We hypothesize that function words may be one indication of the closeness of the relationship between two interlocutors, so this forms the bulk of our transcript-derived features. We extracted the rates of a number of transcribed function words as uttered by the speakers. These are listed in Table 1 along with some examples from each of the classes. We realize that certain

Table 1: *List of function word-derived features.*

Feature Type	Example
Rate of Subject Pronouns	(I, he, we, ...)
Rate of Object Pronouns	(me, her, us, ...)
Rate of Possessive Pronouns	(my, mine, hers, ours, ...)
Rate of Articles	(a, an, the, ...)
Rate of Relative Pronouns	(that, who, which, ...)
Rate of Conjunctions	(and, but, not, ...)
Rate of Prepositions	(to, of, for, with, ...)
Rate of Auxiliary Verbs	(be, have, do, ...)
Rate of Modal Verbs	(can, could, would, ...)
Rate of Negative Words	(not, never, ...)

function words (eg. "that") have multiple interpretations–in this case, acting as any of the classes: pronoun, conjunction, adverb and adjective. We make no distinction among these.

In addition to the rates of words listed in Table 1, we extracted a number of timing and turn-taking features (word rate per second, rate of utterances—words or sounds uninterrupted by the interlocutor, number of times one speaker cuts off another and the delay between speaker turns) as well as the rate of disfluencies. Together, these may belie the amount planning the speakers undergo, which may be indicative of how comfortable or familiar they are with each other. Finally, we extracted the rates of proper nouns and foreign words, which may indicate people or places in common to the speakers.

5. Results

In this section we describe the results of our experiments: 1) classification based on the full conversation, 2) classification based on observation of a single side of the conversation, 3) an analysis of classification performance using less than thirty minutes of speech. In each of these experiments we investigate 4 classifiers from Weka 1) SMO, an SVM optimization algorithm, 2) J48, a decision tree algorithm, 3) Naive Bayes, and 4) BayesNet, a Bayesian Network learning algorithm. All experiments are done using ten-fold cross validation.

5.1. Full Conversation Recognition

First we create feature vectors composed from both sides of the acoustic or transcript-derived features. The transcript features unify the rate and counts and also include any overlaps in speaker turns. Overlaps can be a good indicator of interpersonal stance and relationship. However, we believe this information to be informative when used with acoustic features. We therefore include this with the acoustic features as well. We predict the relationship of the speakers as FRIENDS or FAMILY and provide results from all four classifiers, SMO, J48, Naive Bayes, and BayesNet, on ten-fold cross validation in Table 2.

Table 2: *Results for Full Conversation Recognition.*

Classifier	Acoustic Features	Acoustic + Overlap	Text Features	Acoustic + Text
SMO	42.85%	44.64%	57.14%	39.28%
J48	55.57%	53.57%	57.14%	60.71%
Naive Bayes	44.64%	46.42%	60.71%	57.14%
BayesNet	55.35%	55.35%	55.35%	51.78%

From the table we can see that this is a difficult task and classification is close to chance (50%). While two classifiers, J48 and BayesNet, show over 5% above chance prediction using only acoustic features, the other two perform over 5% below chance. The addition of overlap features accounts for a rise with the underperforming classifiers by nearly 2%, but this harms the performance of the J48 classifier. The text features show more promising results. All the classifiers perform above chance. The best classifier in this case is Naive Bayes with prediction rate of 60.71%. Finally, we combined acoustic and text features. The results are lower compared with text features alone, with the exception of J48 classifier that shows increase in performance.

From the experiments we discovered that MFCC based features and text features that based on counts of words "my", "i", "so", "of", "a", and "had" work the best. The full list of useful features as determined by Information Gain criteria is in Table 3. We find that conversations between friends have a higher rate

Table 3: *The most discriminative features for Full Conversation Classification. Arrows indicate positive or negative correlation with FRIENDS conversations.*

Text Feature	Gain Value	Dir.
my_PER_(TOKEN,UTT,SEC)	0.19, 0.19, 0.19	↑
i_PER_UTT	0.19	↑
so_PER_(TOKEN,UTT,SEC)	0.18, 0.18, 0.16	↑
of_PER_SECOND	0.15	↑
a_PER_SECOND	0.15	↑
had_PER_UTT	0.13	↑
max Δ mfcc[6]	0.18	↑
Linear Regression slope F0	0.18	↓
max Δ mfcc[10]	0.16	↓

of the first-person pronouns, "my" and "i", than conversations between family members. This is somewhat consistent with the observations of Roberts and Dunbar [19] finding that friendly relationships require more maintenance. Conversations among family members contain fewer first-person topics, leaving the conversation to be more focused on other people or possibly more goal-oriented (as discussed by Patrick and Metcalf [2]). We hesitate to draw firm conclusions based on the acoustic features, but we find spectral and pitch differences in these conversations.

5.2. Single Side Recognition

If you were to hear someone's phone conversation, you will most likely hear only one speaker or side of the conversation. In this section, we pose the question: is it possible to distinguish interpersonal roles through examination of a single speaker? Here we investigate how classification performs using acoustic features extracted from only one side of the conversation.

In order to avoid bias introduced from using one person for training and the conversation partner for testing, we work only with the call initiators or call receivers. Speaker A, the caller, is extracted from channel 1 of each conversation and used in cross-validation experiments. Then we repeat the experiments for the call recipient, speaker B, drawn from channel 2. The results can be found at the Table 4.

Table 4: *Results for Single Side Recognition.*

Classifier	Speaker A	Speaker B	Average
SMO	37.5%	48.21%	42.85%
J48	60.71%	62.50%	60.60%
Naive Bayes	50 %	37.50%	44.06%
BayesNet	55.35%	73.21%	64.28%

Interestingly, on balance, the single side classification performance is somewhat better than classification based on the full conversation (cf. Table 2). The results show that three classification methods perform better on the speaker B, while Naive Bayes performs better on speaker A. This phenomena maybe explained in part due to specifics of the data. In the CALLHOME corpus, 85% of the participants placed calls from North America to other countries. In this situation speaker B is more likely to share the experiences and stories about living in a foreign country and provide more salient acoustic information to the classifiers. While the content from the caller is less discriminative of an interpersonal relationship, the call recipient's content varies more clearly based on if they are talking to a friend or family member. As a result, in some cases examining only one side of the conversation is sufficient. Further investigation and comparison of acoustic and text features reveals that speaker B predictions rely on acoustic features more then speaker A. The most predictive features (acoustic and text) and Information Gain values are presented in Table 5.

Table 5: *The most helpful features for speaker A and B. Checkmarks (✓) indicate which task the feature was useful for with associated Information Gain. Arrows indicate positive or negative correlation with FRIENDS conversations*

Feature	Gain Value	A	B
had_PER_(TOKEN,UTT)	0.16, 0.20	✓ ↑	
pause_before_UTT	0.18	✓ ↓	
an_PER_(TOKEN,UTT)	0.11, 0.11	✓ ↑	
my_PER_UTT	0.13	✓ ↑	
may_PER_SEC	0.16	✓ ↑	
his_PER_(TOKEN,UTT,SEC)	0.12		✓ ↑
could_PER_TOKEN	0.11		✓ ↑
with_PER_SECOND	0.095		✓ ↑
max Δ mfcc[(3,10)]	0.16, 0.18	✓ ↑	
Lin. Reg. slope mfcc[10]	0.16	✓ ↓	
min, range Δ mfcc[5]	0.17, 0.30		✓ ↓↑
range mfcc[5]	0.20		✓ ↓
min, minPos mfcc[4]	0.21, 0.21		✓ ↑↓
range, max mfcc[6]	0.20, 0.19		✓ ↓↓
maxPos F0	0.17		✓ ↓
maxPos Δ F0	0.15		✓ ↑

We find that the discriminative features show some differences by speaker side. The use of "my" in FRIENDS conversations seems localized to the call-initiator, while speaker B, the recipient, is more likely to use "his". A turn-taking feature, the length of a pause before each utterance, becomes important in this context. However, in single side analysis, this is a measure of the duration between turns, rather than a measure of smooth turn-taking. This indicates shorter speaker B turns in FRIENDS conversations. Examining the acoustics, we find a set of spectral features to be discriminative. The pitch features indicate an earlier maximum pitch and delta pitch. This may be evidence of more rapid assimilation; this remains a topic of future study.

5.3. Segments of Conversation

People unconsciously speak differently with friends, colleagues, or family. Work by Neiderhoffer and Pennebaker [20] investigated this phenomena, known as assimilation, accommodation or entrainment, in the context of text dialogs. They found that people tend to coordinate word choice and style. In the acoustic domain, Levitan et al. [21] examined the realizations of entrainment in game-playing dyads. Based on these findings, we decided to look into the mid-segments of conversations, where both speakers are involved with the talk. The assumption is that they will have already assimilated to a common speaking style. From each file, we extract acoustic features from a fifteen minute segment starting ten minutes into the dialog. Here we find that despite being drawn from a smaller segment of the conversation, using most classifiers performance actually increases. This suggests that 1) this classification can be reliably performed with shorter observations and 2) that the middle of a conversation may be a reliable locus for analysis.

Table 6: *Results for Start and End of Conversations.*

Classifier	Full	Segment	Delta
SMO	42.85%	48.21%	5.36%
J48	55.57%	57.14%	1.57%
Naive Bayes	44.64%	50.00%	5.36%
BayesNet	55.35%	51.78%	-3.57%

6. Conclusion and Future Work

In this paper, we present preliminary results on the classification of interpersonal roles between conversational partners. Specifically we examine the difference between conversations between friends and family. We have annotated the CALLHOME English corpus with the relationship between participants and made these annotations available. Our initial findings suggest that it is possible to distinguish friends from family on the basis of some low level lexical and acoustic signals. Further work will develop the acoustic analysis to examine evidence of entrainment between partners. The current lexical feature set is based on the usage of common function words. We will develop these to capture more discourse and dialog qualities.

We also compare the ability to perform this analysis on just a single side of a conversation. We find that performance is, on balance, somewhat better. In particular, the participant who receives a telephone call is easier to classify than the conversation initiator. Identifying the differences between these two conversational roles will be a source for future investigations.

7. Acknowledgements

We thank Min Ma, Michelle Morales, Rachel Rakov, Syed Reza, and Clayton Violand for their annotation work.

This material is based on research sponsored by DARPA under agreement number FA8750-13- 2-0041. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

8. References

[1] A. Canavan, D. Graff, and G. Zipperlen, "Callhome american english speech." Linguistic Data Consortium, Philadelphia, 1997.

[2] E. Patrick and C. J. Metcalf, "Mediated communication between extended family and friends: A case study," Motorola Labs, Tech. Rep., 2001.

[3] S. Yaman, D. Hakkani-Tür, and G. Türkhan, "Social role discovery from spoken language using dynamic bayesian networks," in *INTERSPEECH*, 2010, pp. 2870–2873.

[4] R. Barzilay, M. Collins, J. Hirschberg, and S. Whittaker, "The rules behind roles: Identifying speaker role in radio broadcasts," in *AAAI*, 2000.

[5] B. Hutchinson, B. Zhang, and M. Ostendorf, "Unsupervised broadcast conversation speaker role labeling," in *ICASSP*, 2010.

[6] A. Janin, J. Ang, S. Bhagat, R. Dhillon, J. Edwards, J. Macías-guarasa, N. Morgan, B. Peskin, E. Shriberg, A. Stolcke, C. Wooters, and B. Wrede, "The icsi meeting project: Resources and research," in *in Proc. of ICASSP 2004 Meeting Recognition Workshop*. Prentice Hall, 2004.

[7] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus," *Language Resources and Evaluation Journal*, vol. 41, no. 2, pp. 181–190, 2007.

[8] N. Garg and D. Hakkani-Tür, "Speaker role detection in meetings using lexical information and social network analysis," ICSI, Tech. Rep., 2008.

[9] A. Sapru and F. Valente, "Automatic speaker role labeling in ami meetings: Recognition of formal and social roles," in *ICASSP*, 2012.

[10] N. Campbell, "Accounting for voice-quality variation," in *Speech Prosody*, 2004.

[11] G. M. Fitzsimons and A. C. Kay, "Language and interpersonal cognition: Causal effects of variations in pronoun usage on perceptions of closeness," *Personality and Social Psychology Bulletin*, vol. 30, no. 5, pp. 547–557, 2004. [Online]. Available: http://psp.sagepub.com/content/30/5/547.abstract

[12] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg, "Echoes of power: Language effects and power differences in social interaction," in *Proceedings of the 21st International Conference on World Wide Web*, ser. WWW '12. New York, NY, USA: ACM, 2012, pp. 699–708. [Online]. Available: http://doi.acm.org/10.1145/2187836.2187931

[13] M. E. Ireland, R. B. Slatcher, P. W. Eastwick, L. E. Scissors, E. J. Finkel, and J. W. Pennebaker, "Language style matching predicts relationship initiation and stability," *Psychological Science*, vol. 22, no. 1, pp. 39–44, 2011. [Online]. Available: http://pss.sagepub.com/content/22/1/39.abstract

[14] R. Ranganath, D. Jurafsky, and D. A. Mcfarland, "Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates," *Comput. Speech Lang.*, vol. 27, no. 1, pp. 89–115, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.csl.2012.01.005

[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutmann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[16] B. Schuller, S. Steidl, and A. Batliner, "The interspeech 2009 emotion challenge," in *Interspeech*, 2009.

[17] F. Eyben, M. Wöllmer, and B. Schuller, "opensmile - the munich versatile and fast open-source audio feature extractor," in *ACM Multimedia*, 2010.

[18] C. Chung and J. W. Pennebaker, "The psychological functions of function words," *Social communication*, pp. 343–359, 2007.

[19] S. G. B. ROBERTS and R. I. M. DUNBAR, "Communication in social networks: Effects of kinship, network size, and emotional closeness," *Personal Relationships*, vol. 18, no. 3, pp. 439–452, 2011.

[20] K. G. Niederhoffer and J. W. Pennebaker, "Linguistic style matching in social interaction," *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 337–360, 2002.

[21] R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, and A. Nenkova, "Acoustic-prosodic entrainment and social behavior," in *NAACL-HLT*, 2012.