An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation

Mahaveer Jain[1], **John McDonough**[1], **Gahgene Gweon**[2], **Bhiksha Raj**[1], **Carolyn Penstein Rosé**[1,2]
1. Language Technologies Institute; 2. Human Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{mmahavee,johnmcd,ggweon,bhiksha,cprose}@cs.cmu.edu

Abstract

Speech style accommodation refers to shifts in style that are used to achieve strategic goals within interactions. Models of stylistic shift that focus on specific features are limited in terms of the contexts to which they can be applied if the goal of the analysis is to model socially motivated speech style accommodation. In this paper, we present an unsupervised Dynamic Bayesian Model that allows us to model stylistic style accommodation in a way that is agnostic to which specific speech style features will shift in a way that resembles socially motivated stylistic variation. This greatly expands the applicability of the model across contexts. Our hypothesis is that stylistic shifts that occur as a result of social processes are likely to display some consistency over time, and if we leverage this insight in our model,we will achieve a model that better captures inherent structure within speech.

1 Introduction

Sociolinguistic research on speech style and its resulting social interpretation has frequently focused on the ways in which shifts in style are used to achieve strategic goals within interactions, for example the ways in which speakers may adapt their speaking style to suppress differences and accentuate similarities between themselves and their interlocutors in order to build solidarity (Coupland, 2007; Eckert & Rickford, 2001; Sanders, 1987). We refer to this stylistic convergence as speech style accommodation. In the language technologies community, one targeted practical benefit of such modeling has been the achievement of more natural interactions with speech dialogue systems (Levitan et al., 2011).

Monitoring social processes from speech or language data has other practical benefits as well, such as enabling monitoring how beneficial an interaction is for group learning (Ward & Litman, 2007; Gweon, 2011), how equal participation is within a group (DiMicco et al., 2004), or how conducive an environment is for fostering a sense of belonging and identification with a community (Wang et al., 2011).

Typical work on computational models of speech style accommodation have focused on specific aspects of style that may be accommodated, such as the frequency or timing of pauses or backchannels (*i.e.*, words that show attention like 'Un huh' or 'ok'), pitch, or speaking rate (Edlund et al., 2009; Levitan & Hirschberg, 2011). In this paper, we present an unsupervised Dynamic Bayesian Model that allows us to model speech style accommodation in a way that does not require us to specify which linguistic features we are targeting. We explore a space of models defined by two independent factors, namely the direct influence of one speaker's style on another speaker's style and the influence of the relational gestalt between the two speakers that motivates the stylistic accommodation, and thus may keep the accommodation moving consistently, with the same momentum. Prior work has explored the influence of the first factor. However, because accommodation reflects social processes that extend over time within an interaction, one may expect a certain consistency of motion within the stylistic shift. Furthermore, we can leverage this consistency of style shift to identify socially meaningful variation without specifying ahead of time which

Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 787–797,
Avignon, France, April 23 - 27 2012.

particular stylistic elements we are focusing on. Our evaluation provides support for this hypothesis.

When stylistic shifts are focused on specific linguistic features, then measuring the extent of the stylistic accommodation is simple since a speaker's style may be represented on a one or two dimensional space, and movement can then be measured precisely within this space using simple linear functions. However, the rich sociolinguistic literature on speech style accommodation highlights a much greater variety of speech style characteristics that may be associated with social status within an interaction and may thus be beneficial to monitor for stylistic shifts. Unfortunately, within any given context, the linguistic features that have these status associations, which we refer to as *indexicality*, are only a small subset of the linguistic features that are being used in some way. Furthermore, which features carry this indexicality are specific to a context. Thus, separating the socially meaningful variation from variation in linguistic features occurring for other reasons is akin to searching for the proverbial needle in a haystack. It is this technical challenge that we address in this paper.

In the remainder of the paper we review the literature on speech style accommodation both from a sociolinguistic perspective and from a technological perspective in order to motivate our hypothesis and proposed model. We then describe the technical details of our model. Next, we present an experiment in which we test our hypothesis about the nature of speech style accommodation and find statistically significant confirming evidence. We conclude with a discussion of the limitations of our model and directions for ongoing research.

2 Theoretical Framework

Our research goal is to model the structure of speech in a way that allows us to monitor social processes through speech. One common goal of prior work on modeling speech dynamics has been for the purpose of informing the design of more natural spoken dialogue systems (Levitan et al., 2011). The practical goal of our work is to measure the social processes themselves, for example in order to estimate the extent to which group discussions show signs of productive consensus building processes (Gweon, 2011). Much prior work on modeling emotional speech has sought to identify features that themselves have a social interpretation, such as features that predict emotional states like uncertainty (Liscombe et al., 2005), or surprise (Ang et al., 2002), or social strategies like flirting (Ranganath et al., 2009). However, our goal is to monitor social processes that evolve over time and are reflected in the change in speech dynamics. Examples include fostering trust, forming attachments, or building solidarity.

2.1 Defining Speech Style Accommmodation

The concept of what we refer to as Speech Style Accommodation has its roots in the field of the Social Psychology of Language, where the many ways in which social processes are reflected through language, and conversely, how language influences social processes, are the objects of investigation (Giles & Coupland, 1991). As a first step towards leveraging this broad range of language processes, we refer to one very specific topic, which has been referred to as entrainment, priming, accommodation, or adaptation in other computational work (Levitan & Hirschberg, 2011). Specifically we refer to the finding that conversational partners may shift their speaking style within the interaction, either becoming more similar or less similar to one another.

Our usage of the term accommodation specifically refers to the process of speech style convergence within an interaction. Stylistic shifts may occur at a variety of levels of speech or language representation. For example, much of the early work on speech style accommodation focused on regional dialect variation, and specifically on aspects of pronunciation, such as the occurrence of post-vocalic "r" in New York City, that reflected differences in age, regional identification, and socioeconomic status (Labov, 2010a,b). Distribution of backchannels and pauses have also been the target of prior work on accommodation (Levitan & Hirschberg, 2011). These effects may be moderated by other social factors. For example, Bilous & Krauss (1988) found that females accommodated to their male partners in conversation in terms of average number of words uttered per turn. For example, Hecht et al. (1989) reported that extroverts are more listener adaptive than introverts and hence extroverts converged more in their data.

Accommodation could be measured either from textual or speech content of a conversation. The former relates to "what" people say whereas the latter to 'how' they say it. We are only interested in measuring accommodation from speech in this work. There has been work on convergence in text such as syntactic adaptation (Reitter et al., 2006) and language similarity in online communities (Huffaker et al., 2006).

2.2 Social Interpretation of Speech Style Accommodation

It has long been established that while some speech style shifts are subconscious, speakers may also choose to adapt their way of speaking in order to achieve social effects within an interaction (Sanders, 1987). One of the main motives for accommodation is to decrease social distance. On a variety of levels, speech style accommodation has been found to affect the impression that speakers give within an interaction. For example, Welkowitz & Feldstein (1970) found that when speakers become more similar to their partners, they are liked more by partners. Another study by Putman & Street Jr (1984) demonstrated that interviewees who converge to the speaking rate and response latency of their interviewers are rated more favorably by the interviewers. Giles et al. (1987) found that more accommodating speakers were rated as more intelligent and supportive by their partners. Conversely, social factors in an interaction affect the extent to which speakers engage in, and some times chose not to engage in, accommodation. For example, Purcell (1984) found that Hawaiian children exhibit more convergence in interactions with peer groups that they like more. Bourhis & Giles (1977) found that Welsh speakers while answering to an English surveyor broadened their Welsh accent when their ethnic identity was challenged. Scotton (1985) found that few people hesitated to repeat lexical patterns of their partners to maintain integrity. Nenkova et al. (2008) found that accommodation on high frequency words correlates with naturalness, task success, and coordinated turn-taking behavior.

2.3 Computational models of speech style accommodation

Prior research has attempted to quantify accommodation computationally by measuring similarity of speech and lexical features either over full conversations or by comparing the similarity in the first half and the second half of the conversation. For example, Edlund et al. (2009) measure accommodation in pause and gap length using measures such as synchrony and convergence. Levitan & Hirschberg (2011) found that accommodation is also found in special social behaviors within conversation such as backchannels. They show that speakers in conversation tend to use similar kinds of speech cues such as high pitch at the end of utterance to invite a backchannel from their partner. In order to measure accommodation on these cues, they compute the correlation between the numerical values of these cues used by partners.

In our work we measure accommodation using Dynamic Bayesian Networks (DBNs). Our models are learnt in an unsupervised fashion. What we are specifically interested in is the manner in which the influence of one partner on the other is modeled. What is novel in our approach is the introduction of the concept of an accommodation state, or relational gestalt variable, which essentially models the momentum of the influence that one partner is having on the other partner's speaking style. It allows us to represent structurally the insight that accommodation occurs over time as a reflection of a social process, and thus has some consistency in the nature of the accommodation within some span of time. The prior work described in this section can be thought of as taking the influence of the partner's style directly on the speaker's style within an instant as the floor shifts from one speaker to the next. Thus, no consistency in the manner in which the accommodation is occurring is explicitly encouraged by the model. The major advantage of consistency of motion within the style shift over time is that it provides a sign post for identifying which style variation within the speech is salient with respect to social interpretation within a specific interaction so that the model may remain agnostic and may thus be applied to a variety of interactions that differ with respect to which stylistic features are salient in this respect.

3 A Dynamic Bayesian Network Model for Conversation

Speech stylistic information is reflected in prosodic features such as pitch, energy, speak-

ing rate etc. In this work, we leverage on several of these speech features to quantify accommodation. We propose a series of models that can be trained unsupervised from speech features and can be used for predicting accommodation. The models attempt to capture the dependence of speech features on speaking style, as well as the effect of persistence and accommodation on style. We use a dynamic Bayesian network (DBN) formalism to capture these relationships. Below we briefly review DBNs, and subsequently describe the speech features used, and the proposed models.

3.1 Dynamic Bayesian Networks

The theory of Bayesian networks is well documented and understood (Jensen, 1996; Pearl, 1988). A Bayesian network is a probabilistic model that represents statistical relationships between random variables via a directed acyclic graph (DAG). Formally, it is a directed acyclic graph whose nodes represent random variables (which may be observable quantities, *latent* unobservable variables, or hypotheses to be estimated). Edges represent *conditional* dependencies; nodes which are connected by an edge represent random variables that have a direct influence on one another. The entire network represents the joint probability of all the variables represented by the nodes, with appropriate factoring of the conditional dependencies between variables.

Consider, for instance, a joint distribution over a set of random variables $x_1, x_2, \cdots, x_n$, modeled by a Bayesian network. Let $\mathcal{V} = v_1, v_2, \cdots, v_n$ represent the set of n nodes in the network, representing the random variables $x_1, x_2, \cdots, x_n$ respectively. Let $\wp(v_i)$ represent the set of parent nodes of v_i, *i.e.* nodes in $\mathcal{V}$ that have a directed edge into a node v_i. Then, by the dependencies specified by the network, $P(x_i|x_1, x_2, \cdots, x_n) = P(x_i|x_j : v_j \in \wp(v_i))$. In other words, any variable x_i is directly dependent only on its parent variables, *i.e.* the random variables represented by the nodes in $\wp(v_i)$, and is independent of all other variables given these variables. The joint probability of $x_1, x_2, \cdots, x_n$ is hence given by

$$p(x_1, x_2, ..., x_n) = \prod_i p(x_i|x_{\pi_i}) \qquad (1)$$

Where x_{π_i} represents $\{x_j : v_j \in \wp(v_i)$, *i.e.* the parents of x_i in the network. We note that not all of these variables need to be observable; often in such models several of the variables are unobservable, *i.e.* they are *latent*. In order to obtain the joint distribution of the observable variables the latent variables must be marginalized out. *I.e.* if $x_1, \cdots, x_m$ are observable and $x_{m+1}, \cdots, x_n$ are latent, $P(x_1, \cdots, x_m) = \sum_{x_{m+1}, \cdots, x_n} P(x_1, x_2, \cdots, x_n)$.

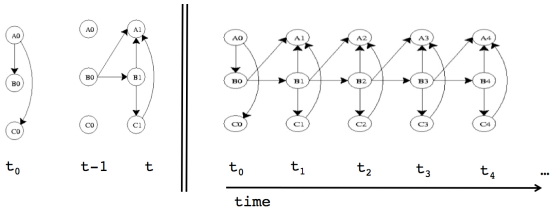

Figure 1: An example Dynamic Bayesian Network (DBN) showing the temporal relationship between three random variables (A,B and C). A is observered and dependent on two hidden variables B and C. Directed edges across time ($t-1 \rightarrow t$) indicate temporal relationships between variables. In this example, the variables A_t and B_t are both dependent on B_{t-1} with the relationship defined through conditional distributions $P(A_t|B_{t-1})$ and $P(B_t|B_{t-1})$.

Dynamic Bayesian networks (DBNs) further represent time-series data through a recurrent formulation of a basic Bayesian network that represents the relationship between variables. Within a DBN a set of random variables at each time instance t is represented as a static Bayesian Network with temporal dependencies to variables at other instants. Namely, the distribution of a variable $x_{i,t}$ at time t is dependent on other variables at times $t-\tau$, $x_{j,t-\tau}$ through conditional probabilities of the form $Pr(x_{i,t}|x_{j,t-\tau})$. An example DBN, consisting of three variables (A, B and C), two of which have temporal dependencies is shown in Figure 1.

One benefit of the DBN formalism is that in addition to providing a compact graphical way of representing statistical relationships between variables in a process, the constrained, directed network structure also allows for simplified inference. Moreover, the conditional distributions associated with the network are often assumed not to vary over time, i.e. $Pr(x_{i,t}|x_{j,t-\tau}) = Pr(x_{i,t'}|x_{j,t'-\tau})$. This allows for a very compact representation of DBNs and allows for efficient Expectation-Maximization (EM) learning algorithms to be applied.

In the discussion that follows we do not explicitly specify the random variables and the form of the associated probability distributions, but only present them graphically. The joint distribution of the variables should nevertheless be obvious from the figures. We employ EM to learn the parameters of the models from training data, and the junction tree algorithm (Lauritzen & Spiegelhalter, 1988) to perform inference.

3.2 Speech Features

We characterize conversations as a series of spoken turns by the partners. We characterize the speech in each turn through a vector that captures several aspects of the signal that are salient to style. We used the OPENSmile toolkit (opensmile, 2011) to compute the features. Specifically, within each turn the speech was segmented into analysis windows of 50ms, where adjacent windows overlapped by 40ms. From each analysis window a total of 7 features were computed: voice probability, harmonic to noise ratio, voice quality , three measures of pitch (F_0, F_0^{raw}, F_0^{env}), and loudness. A 10-bin histogram of feature values was computed for each of these features, which was then normalized to sum to 1.0. The normalized histogram effectively represents both the values and the fluctuation in the features. For instance, a histogram of loudness values captures the variation in the loudness of the speaker within a turn. The logarithms of the normalized 10-bin histograms for the 7 features were concatenated to result in a single 70-dimensional observation vector for the turn. These 70 dimensional observation vectors for each turn of any speaker are represented in our model as o_t^i where t is turn index and i is speaker index.

3.3 Elements of the Models

In this section we formally describe the elements of our model.

Speaking Style State: These states represent the speaking styles of the partners in a conversation. We represent these states as s_t^i, where t represent turn index and i represents speaker index. These states are assumed to belong to a finite, discrete set $\mathcal{S} = \{s_1, s_2, \cdots, s_k\}$, *i.e.* $s_t^i \in \mathcal{S}\ \forall(i,t)$.

Accommodation State: An accommodation state represents the indirect influence of partners on each other in a conversation. In our present design, it can take a value of either 1 or 0. These states are represented as A_t, where t is turn index.

Observation Vector: The observation vectors are the feature vectors o_t^i computed for each turn.

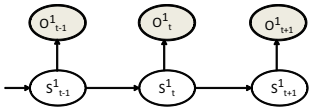

Figure 2: The basic generative model.

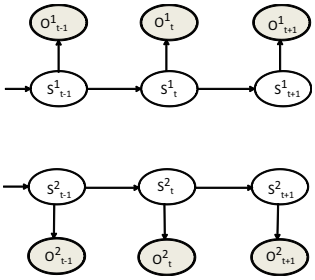

Figure 3: ISM: The dynamics of each speaker are independent of the other speaker.

3.4 Models for Accommodation

Our models embody two premises. First, a person's speech in any turn is a function of his/her speaking style in that turn. Second, a person's speaking style at any turn depends not only by their own personal biases, but also by their accommodation to their partner. We represent these dependencies as a DBN.

Our basic model to represent the generation of speech (i.e. speech features) by a speaker in the absence of other influences is shown in Figure 2. The speech features o_t^i in any turn depend only on the speaking style s_t^i in that turn. The style s_t^i in any turn depends on the style s_{t-1}^i in the previous turn, to capture the speaker-specific patterns of variation in speaking style. We note that this is a rather simple model and patterns of variation in style are captured only through the statistical dependence between styles in consequent turns.

We now build our models for accommodation on this basic model.

3.4.1 Style-based models

Our two first models assume that accommodation is demonstrated as a direct dependence of a person's speaking sytle on their partner's style. Therefore the models only consider speaking styles.

The Independent Speaker Model

Our simplest model for a conversation assumes

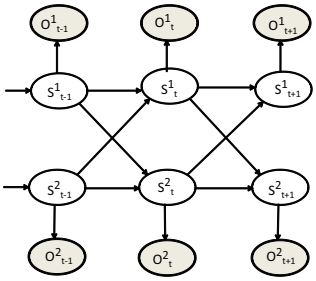

Figure 4: CSDM: A speaker's style depends on their partner's style at the previous turn.

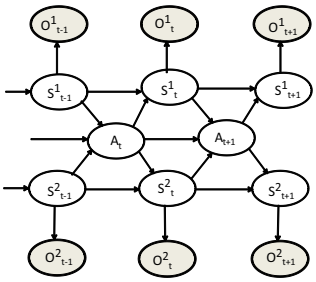

Figure 5: SASM: Both partners' styles depend on mutual accommodation to one another.

that each person's speaking style evolves independently, uninfluenced by their partner. The DBN for this is shown in Figure 3. We refer to this model as the *Independent Speaker Model* (ISM). Note that the set of values that the style states can take is common for both speakers. The speaking styles for the two speakers may be said to be confluent in any turn if both of them are in the same style state at that turn.

The Cross-speaker Dependence Model

Intuitively, in a conversation speakers are influenced by their partners' speaking style in previous turns. The *Cross-Speaker Dependence Model* (CSDM) represents this dependence as shown in the DBN in Figure 4. In this model a person's speaking style depends on both their own and their partner's speaking styles in the previous turn.

3.4.2 Accommodation state models

Accommodation state models assume that conversations actually have an underlying state of accommodation, and that speakers in fact vary their speaking styles in response to it. We models this through a binary-valued accommodation state that is embedded into the DBN. We posit two types of accommodation state models.

The Symmetric Accommodation State Model

In the *symmetric* accommodation state model

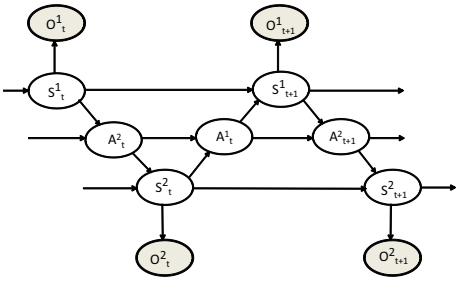

Figure 6: AASM: Accommodation state associated with every speaker turn

(SASM) we assume that accommodation is a jointly experienced characteristic of the conversation at any time, which enjoys some persistence, but is also affected by the speaking styles exhibited by the speakers at each turn. The accommodation at any time in turn affects the speaking styles of both speakers in the next turn. The DBN for this model is shown in Figure 5.

The Asymmetric Accommodation State Model

The *asymmetric* accommodation state model (AASM) represents accommodation as a speaker-turn-specific characteristic. In any turn, the accommodation for a speaker depends chiefly on their partner's most recent speaking style. The accommodation state can change after each speaker turn. Figure 6 shows the DBN for this model. Note that this model captures the asymmetric nature of accommodation, e.g. it may be the case that only one of the speakers is accommodating. For instance, if if $a^1_t = 0$ and $a^2_t = 1$, only speaker2 is accommodating but not speaker1.

3.4.3 Accommodated style dependence models

While accommodation state models explicitly models accommodation, they do not explicitly represent how it is expressed. In reality, accommodation is a process of convergence – an accommodating speaker's speaking style may be expected to converge toward that of their partner. In other words, the person's speaking style depends not only on whether they are accommodating or not, but also on their partner's style at the previous turn. *Accommodated style dependence models* explicitly represent this dependence.

The Symmetric Accommodated Style Dependence Model

The *Symmetric Accommodated Style Dependence Model* (SASDM) extends the SASM, to in-

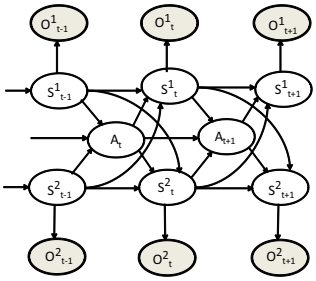

Figure 7: SASDM: A speaker's style depends both on mutual accommodation and the partner's style in the previous turn.

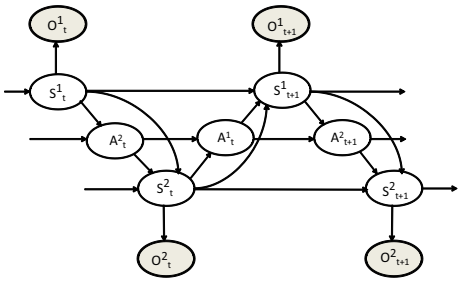

Figure 8: AASDM: The accommodation state associated with every speaker and a speaker's style depends on the partner's style.

dicate that a speaker's style in any turn depends both on accommodation and on their partner's style in the previous turn. Figure 7 shows the DBN for this model.

Asymmetric Accommodated Style Dependence Model

The *Asymmetric Accommodated Style Dependence Model* (AASDM) extends the AASM by adding a direct dependence between a speaker's style and their partner's style in their most recent turn. The DBN for this is shown in Figure 8.

3.5 Interpreting the states

We note that we have referred to the states in the models above as "style" states. In reality, in all cases, we learn the parameters of the model in an unsupervised manner, since the data we use to train it do not have either speaking style or accommodation indicated (although, if they were labeled, the labels could be employed within our models). Consequently, we have no assurance that the states learned will actually correspond to speaking styles. They can only be considered a *proxy* for speaking style. Nevertheless, if both speakers are in the same state, they can both be expected to be producing similar prosodic features, as represented in the observation vectors. It is hence reasonable to assume that they are both speaking in similar style. Similarly, the accommodation state cannot be expected to actually depict accommodation; nevertheless, it can capture the dependencies that govern when the two speakers are likely to be in the same state.

4 Evaluation

The model we have just described allows us to investigate two separate aspects of our concept of speech style accommodation. The first aspect is that style accommodation occurs as a local influence of one speaker's style on the other speaker's style, as depicted by direct links between style states. The second aspect is that although this is a local phenomenon, because it is a reflection of a social process that extends over a period of time, there will be some persistence of accommodation over longer periods of time, as characterized by the accommodation state. We presented two different operationalizations of the accommodation state above, namely Asymmetric and Symmetric.

Accommodation is a phenomenon that occurs within interactions between speakers; we can expect not to observe accommodation occurring between individuals that have never met and are not interacting. On average, then, we expect to see more evidence of speech style accommodation in pairs of individuals who are interacting (i.e., Real Pairs) than in pairs of individuals who are not interacting and have never met (i.e., Constructed Pairs). Thus, we may evaluate the extent to which our model is sensitive to social dynamics within pairs by the extent to which it is able to distinguish between true conversation between Real Pairs of speaker and synthetic conversation between Constructed Pairs. A similar experimental paradigm has been adopted in prior work on speech style accommodation (Levitan et al., 2011).

Hypothesis: Our hypothesis is that models that explicitly represent the notion that accommodation occurs over a span of time with consistency of momentum will achieve better success at distinguishing between Real Pairs and Constructed Pairs than models that do not.

Experimental Manipulation: Thus, using the model we have just described, we are able to test our hypothesis using a 2×3 factorial design in which one factor is the inclusion of direct links from the style of one speaker to the style

of the other speaker, which we refer to as the DirectInfluence (DI) factor, with values True (T) and False (F), and the second factor is the inclusion of links from style states to and from Accommodation states, which we refer to as the IndirectInfluence (II) factor, with values False (F), Asymmetric (A), and Symmetric (S). The result of this 2×3 factorial design are the 6 different models described in Section 3, namely ISM (DI=False, II=False), CSDM (DI=True, II=False), SASM (DI=False, II=Symmetric), AASM (DI=False, II=Asymmetric), SASDM (DI=True, II=Symmetric), and AASDM (DI=True, II= Asymmetric).

Corpus: The success criterion in our experiment is the extent to which models of speech style accommodation are able to distinguish between Real Pairs and Constructed pairs. In order to set up this comparison, we began with a corpus of debates between students about the reasons for the fall of the Ottoman Empire. We obtained this corpus from researchers who originally collected it to investigate issues related to learning from conversational interactions (Nokes et al., 2010). The full corpus contains interactions between 76 pairs of students who interacted for 8 minutes. Within each pair, one student was assigned the role of arguing that the fall of the Ottoman empire was due to internal causes, whereas the other student was assigned the role of arguing that the fall of the Ottoman empire was due to external causes. Each student was given a 4 page packet of supporting information for their side of the debate to draw from in the interaction.

The speech from each participant was recorded on a separate channel. As a first step, we aligned the speech recordings automatically to their transcriptions at the word and turn level. After aligning the corpus at the word level, we identify the turn interval of each partner in the conversation. We use 66 of the debates out of the complete set of 76 for the experiments discussed in this paper. We had to eliminate 10 dialogues where the segmentation and alignment failed. For each of our models, we used the same 3 fold cross-validation.

Participants: Participants were all male undergraduate students between the ages of 18 and 25. In prior studies, it has been shown that accommodation varies based on gender, age and familiarity between partners. This corpus is particularly appropriate because it controls for most of these factors. Furthermore, because the participants did not know each other before the debate, we can assume that if accommodation happened, it was only during the conversation.

Real versus Constructed Pairs: In our analysis below, we compare measured accommodation between pairs of humans who had a real conversation and a constructed pair in which one person from that conversation is paired with a constructed partner, where the partner's side of the conversation was constructed from turns that occurred in other conversations. We set up this comparison in order to isolate speech style convergence from lexical convergence when we evaluate the performance of our model. The difference between the measured accommodation between real and constructed pairs is treated as a weak operationalization of model accuracy at measuring speech style accommodation.

For each of the 20 Real pairs in the test corpus we composed one Constructed Pair. Each Constructed Pair comprised one student from the corresponding Real Pair (i.e., the Real Student) and a Constructed Partner that resembled the real partner in content but not necessarily style. We did this by iterating through the real partner's turns, replacing each with a turn that matched as well as possible in terms of lexical content but came from a different conversation. Lexical content match was measured in terms of cosine similarity. Turns were selected from the other Real pairs. Thus, the Constructed Partner had similar content to the corresponding real partner on a turn by turn basis, but the style of expression could not be influenced by the Real Student. Thus, ideally we should not see evidence of speech style accommodation within the Constructed Pairs.

Experimental Procedure: For each of the four models we computed an Accommodation Score for each of the Real Pairs and Constructed Pairs. In order to obtain a measure that can be used to compute accommodation for all the models considered, we compute the accommodation value as the fraction of turns in a session where partners exhibited the same speaking style.

Results: In order to test our hypothesis we constructed an ANOVA model with Accommodation Score as the dependent variable and DirectInfluence, IndirectInfluence, RealVsConstructed as independent variables. Additionally we included the interaction terms between all pairs of inde-

	DI	II	Real $\mu(\sigma)$	Constructed $\mu(\sigma)$
SASDM	T	S	.54 (.23)	.44 (.29)
SASM	F	S	.54 (.23)	.44 (.29)
CSDM	T	F	.6 (.26)	.52 (.3)
ISM	F	F	.56 (.25)	.51 (.32)
AASM	F	A	.6 (.24)	.51 (.3)
AASDM	T	A	.61 (.24)	.48 (.3)

Table 1: Accommodation measured using different models. Legend: μ=mean, σ = standard deviation, DI = "Direct Influence", II = "Indirect Influence".

pendent variables. Using this ANOVA model, we find a highly significant main effect of the RealVsConstructed factor that demonstrates the general ability of the models to achieve separation between Real Pairs and Constructed Pairs; on average $F(1,780) = 18.22$, $p < .0001$.

However, when we look more closely, we find that although the trend is consistently to find more evidence of speech style accommodation in Real Pairs than in Constructed Pairs, we see differentiation among the models in terms of their ability to achieve this separation. When we examine the two way interactions between DirectInfluence and RealVsConstructed as well as between IndirectInfluence and RealVsConstructed, although we do not find significant interactions, we do find some suggestive patterns when we do the student T posthoc analysis. In particular, when we explore just the interaction between IndirectInfluence links, we find a significant separation between Real vs Constructed pairs for models with Accommodation states, but not for the cases where no Accommodation states are included. However, when we do the same for the interaction between DirectInfluence links and RealVsConstructed, we find significant separation with or without those links. This suggests that IndirectInfluence links are more important than DirectInfluence links. At a finer-grained level, when we examine the models individually, we only find a significant separation between Real and Constructed pairs with the model that includes both DirectInfluence and Symmetric IndirectInfluence links. These results suggest that Symmetric IndirectInfluence links may be slightly better than Asymmetric ones, and that combining DirectInfluence links and Symmetric IndirectInfluence links may be the best combination.

Based on this analysis, we find support for our hypothesis. We find that the model that includes Symmetric IndirectInfluence links and DirectInfluence links is the best balance between representational power and simplicity. The support for the inclusion of DirectInfluence links in the model is weaker than that of IndirectInfluence links, however. On a larger dataset, we may have observed stronger effects of both factors. Even on this small dataset, we find evidence that adding that structure improves the performance of the model without leading to overfitting.

5 Conclusions and Current Directions

In this paper we presented an unsupervised dynamic Bayesian modeling approach to modeling speech style accommodation in face-to-face interactions. Our model was motivated by the idea that because accommodation reflects social processes that extend over time within an interaction, one may expect a certain consistency of motion within the stylistic shift. Our evaluation demonstrated a statistically significant advantage for the models that embodied this idea.

An important motivation for our modeling approach was that it allows us to avoid targeting specific linguistic style features in our measure of accommodation. However, in our evaluation, we only tested our approach on conversations between male undergraduate students discussing the fall of the Ottoman Empire. Thus, while our evaluation provides evidence that we have taken a first important step towards our ultimate goal, we cannot yet claim that we have a model that performs equally effectively across contexts. In our future work, we plan to formally test the extent to which this allows us to accurately measure accommodation within contexts in which very different stylistic elements carry strategic social value.

Another important direction of our current research is to explore how measures of speech style accommodation may predict other important measures such as how positively partners view one another, how successful partners perform tasks together, or how well students learn together.

6 Acknowledgments

We gratefully acknowledge John Levine and Timothy Nokes for sharing their data with us. This work was funded by NSF SBE 0836012.

References

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., & Stolcke, A. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proc. ICSLP*, volume 3, pages 2037–2040. Citeseer.

Bilous, F. & Krauss, R. (1988). Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Language and Communication*, **8**(3), 4.

Bourhis, R. & Giles, H. (1977). The language of intergroup distinctiveness. *Language, ethnicity and intergroup relations*, **13**, 119.

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge Univ Pr.

DiMicco, J., Pandolfo, A., & Bender, W. (2004). Influencing group participation with a shared display. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 614–623. ACM.

Eckert, P. & Rickford, J. (2001). *Style and sociolinguistic variation*. Cambridge Univ Pr.

Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In *Proc. Interspeech*.

Giles, H. & Coupland, N. (1991). *Language: Contexts and consequences.* Thomson Brooks/Cole Publishing Co.

Giles, H., Mulac, A., Bradac, J., & Johnson, P. (1987). Speech accommodation theory: The next decade and beyond. *Communication yearbook*, **10**, 13–48.

Gweon, G. A. P. U. M. R. B. R. C. P. (2011). The automatic assessment of knowledge integration processes in project teams. In *Proceedings of Computer Supported Collaborative Learning*.

Hecht, M., Boster, F., & LaMer, S. (1989). The effect of extroversion and differentiation on listener-adapted communication. *Communication Reports*, **2**(1), 1–8.

Huffaker, D., Jorgensen, J., Iacobelli, F., Tepper, P., & Cassell, J. (2006). Computational measures for language similarity across time in online communities. In *In ACTS: Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech*, pages 15–22.

Jensen, F. V. (1996). *An introduction to Bayesian networks*. UCL Press.

Labov, W. (2010a). *Principles of linguistic change: Internal factors*, volume 1. Wiley-Blackwell.

Labov, W. (2010b). *Principles of linguistic change: Social factors*, volume 2. Wiley-Blackwell.

Lauritzen, S. L. & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, **50**, 157–224.

Levitan, R. & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech*.

Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 113–117. Association for Computational Linguistics.

Liscombe, J., Hirschberg, J., & Venditti, J. (2005). Detecting certainness in spoken tutorial dialogues. In *Proceedings of INTERSPEECH*, pages 1837–1840. Citeseer.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *In Proceedings of ACL-08: HLT. Association for Computational Linguistics*.

opensmile (2011). http://opensmile.sourceforge.net/.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Purcell, A. (1984). Code shifting hawaiian style: childrens accommodation along a decreolizing continuum. *International Journal of the Sociology of Language*, **1984**(46), 71–86.

Putman, W. & Street Jr, R. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, **1984**(46), 97–114.

Ranganath, R., Jurafsky, D., & McFarland, D. (2009). It's not you, it's me: detecting flirting and its misperception in speed-dates. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 334–342. Association for Computational Linguistics.

Reitter, D., Keller, F., & Moore, J. D. (2006). Computational modelling of structural priming in dialogue. In *In Proc. Human Language Technology conference - North American chapter of the Association for Computational Linguistics annual mtg*, pages 121–124.

Sanders, R. (1987). *Cognitive foundations of calculated speech*. State University of New York Press.

Scotton, C. (1985). What the heck, sir: Style shifting and lexical colouring as features of powerful lan-

guage. *Sequence and pattern in communicative behaviour*, pages 103–119.

Wang, Y., Kraut, R., & Levine, J. (2011). To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM conference on computer-supported cooperative work*. ACM.

Ward, A. & Litman, D. (2007). Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education*. Citeseer.

Welkowitz, J. & Feldstein, S. (1970). Relation of experimentally manipulated interpersonal perception and psychological differentiation to the temporal patterning of conversation. In *Proceedings of the 78th Annual Convention of the American Psychological Association*, volume 5, pages 387–388.