Investigating fine temporal dynamics of prosodic and lexical accommodation

Francesca Bonin[1], Céline De Looze[1], Sucheta Ghosh[1], Emer Gilmartin[1], Carl Vogel[1], Anna Polychroniou[2], Hugues Salamin[2], Alessandro Vinciarelli[23], Nick Campbell[1]

[1]Trinity College Dublin, Dublin 2, Ireland
[2] University of Glasgow, School of Computing Science, G128QQ Glasgow, UK
[3] Idiap Research Institute, CP592, 1920 Martigny, Switzerland

boninf,deloozec,ghoshs,gilmare,vogel,nick@tcd.ie, vincia,hsalamin,annap@dcs.gla.ac.uk

Abstract

Conversational interaction is a dynamic activity in which participants engage in the construction of meaning and in establishing and maintaining social relationships. Lexical and prosodic accommodation have been observed in many studies as contributing importantly to these dimensions of social interaction. However, while previous works have considered accommodation mechanisms at global levels (for whole conversations, halves and thirds of conversations), this work investigates their evolution through repeated analysis at time intervals of increasing granularity to analyze the dynamics of alignment in a spoken language corpus. Results show that the levels of both prosodic and lexical accommodation fluctuate several times over the course of a conversation.

Index Terms: prosodic accommodation, lexical alignment, dynamics, task-based interactions

1. Introduction

In spoken interaction, participants have been observed to adapt their speech production to that of their interlocutor. Interspeaker adaptation has been termed accommodation, alignment, convergence, priming as well as synchrony, and has been reported in terms of pronunciation [1, 2], prosody [3, 4, 5, 6], lexicon [7, 8, 9] and syntax [10, 11]. Herein, we will use the term accommodation when the features of a speaker's production change as a function of their partners'. Theories accounting for some of these phenomena include Garrod and Pickering's Interactive Alignment Theory [8] as well as Giles and Coupland's CAT theory [12].

Accommodation mechanisms are a particularly important aspect of spoken interaction as they facilitate, through the alignment of cognitive representations, comprehension and understanding between interlocutors. They correlate with the communicative success of the interaction, by decreasing misunderstandings and attaining goals faster [13, 8, 14]. In addition, accommodation contributes to the social success of the interaction by building rapport (i.e. harmonious relationships and mutual attention) and affiliation [15, 16, 17, 18].

In terms of prosody, accommodation in pitch, (measured as meanf_0 and sdf_0), in voice intensity (measured as mean and sd Intensity), in speech rate (measured in terms of articulation rate and the number and duration of pauses) and vocal activity rhythm has been observed in interviews [3], task-based dialogues and unconstrained conversations [19, 6, 20]. Lexical-syntactical accommodation has also been reported in many works. Earlier studies have measured it as primed words and unprimed words frequency in halves of conversations [21]. Other approaches have investigated syntactical and prosodic alignment using a look-ahead method on a segment of conversation and applied this paradigm in predicting success in dialogues [22], and in leaning outcomes in tutoring scenarios [23]. Relations have further been hypothesized between alignment fluctuation and changes of topic [24].

Many of the above mentioned studies have measured accommodation evolution by comparing mean accommodation levels in large segments of conversations, such as halves or thirds, and have reported increased accommodation over time. However, what makes a conversation an interactive dialogue are the dynamic changes involved in spoken interaction. The interlocutors' roles do not remain static over the whole course of a conversation; as they may change from being inactive to talking, going through phases such as listening, thinking, arguing a point or giving feedback. It can thus be assumed that accommodation undergoes similar dynamic changes, as some earlier and recent work have shown [3, 25, 4, 19, 20, 6]. While lexical and prosodic accommodation mechanisms have been studied at global levels, their evolution or local fluctuation over time is still poorly understood. In this preliminary work, we propose a novel methodology to investigate fluctuations in prosodic and lexical alignment at finer levels of granularity. We investigate the evolution of prosodic (***Experiment 1***) and lexical (***Experiment 2***) accommodation and (2) their alignment (***Experiment 3***) over the course of the conversation. In addition, we explore (3) their role in expressing inter-speaker agreement, defined herein as speakers' consensus on a particular item (***Experiment 4***).

2. Experiments

2.1. Corpus

The data used in the experiments consists of 5 phone calls between unacquainted native (Scottish) English speakers (10 subjects in total).[1] The participants were asked to accomplish the "Winter Survival Task" (WST). They had to discuss whether 12 items, found on the site of a plane crash, could be useful or not for the survival of the passengers. Before the phone call, each participant performed the task individually. The goal of the phone call was to reach consensus in case of disagreement. The only constraint was to discuss the items sequentially. The subjects were provided with Nokia N900 mobile phones and were recorded using the N900 microphones at 44.8 kHz. The five phone calls used in this work were selected out of a larger corpus (60 calls, 120 subjects) based on duration (record-

[1]Conversations consist of about 12 minutes each and 140 turns on average.

ings closer to the average were retained for the experiments) and gender composition (both female in two cases, both men in the other three). The audio of each phone call was manually segmented into speaker turns, and intervals of time corresponding to the decision process about one of the items ("steel wool", "axe", "pistol", "butter", "newspaper", "lighter", "clothing", "canvas", "airmap", "whisky", "compass" and "chocolate"). The five phone calls were also manually transcribed for lexical alignment measurements, using Praat. Transcriptions have been made at the inter-pause level, in two steps (actual transcription and review) in British spelling and using reduced lexical forms. Punctuation has not been transcribed. The informality of the conversation, coupled with the requirement for participants to reach consensus on a number of items, should produce several instances of accommodation.

2.2. Prosodic accommodation measurements

2.2.1. Prosodic cues extraction

Measuring prosodic accommodation necessitates specific requirements as speakers' speech is not aligned in conversations. Speakers talk one after the other and may not accommodate to each other immediately due to the inherent temporally reactive nature of conversational speech. Current approaches comprise two types of methods to measure inter-personal prosodic accommodation: utterance / turn-level-based (e.g. [25]) and time aligned moving average (TAMA) methods [26]. Figure 1 shows the differences between the two approaches. The dotted rectangle (analysis window) refers to the analyzed audio snippets. The third approach, a HYBRID utterance-sensitive approach, presented below, is a trade off between the two extremes.

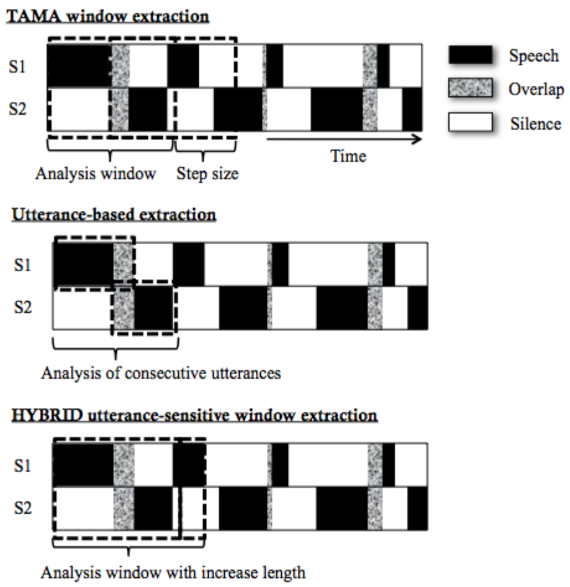

Figure 1: Illustrating comparison of the three analysis methods, i.e. time aligned moving average (TAMA) based, utterance-based, and utterance-sensitive window based or HYBRID.

The turn-level based approaches analyse prosody within speakers' consecutive turn-levels. Such a fine-grained method is interesting as it takes into account speakers' vocal activity rhythm. It however includes the assumption that it is a local phenomenon only. The effect of a speaker's speech characteristics on his/her partners may well be found after some temporal delay, which may exceed the utterance or the turn domain. In contrast to this method, the TAMA (Time-Aligned Moving Average) method proposed by Kousidis [26] analyses the audio from a series of overlapping fixed length windows, averaging values out over the duration of each window. Large and overlapped windows give a smoothed contour for the prosodic parameter being analyzed, while short frames detect more abrupt modifications [26]. While this technique allows for capturing accommodation delays, it cuts utterances randomly, even if a speaker has not yet finished talking. This issue can be resolved through the use of a HYBRID approach that is sensitive to utterance boundaries. In a recent paper [27], we have proposed a HYBRID method based on utterance-based and TAMA methods. Instead of randomly cutting the speech of the speakers, the moving windows are extended to the start and end of the utterances at the left and right boundaries of the window. In particular, this means that average values of prosodic cues are automatically extracted from a series of overlapping windows (frames) of a default-fixed length which are extended to the utterance temporal span at the window boundaries. Such a method therefore allows both the consideration of speakers' vocal activity rhythm and speaker-time-aligned prosodic cue extraction. The argument for an utterance-sensitive model is that the functional aspects of prosodic accommodation may not change within an utterance but rather between utterances. The prosodic features, extracted from the entire utterance, are therefore representative of the utterance prosody and its functions in the interaction. Furthermore, this facilitates the alignment of the dynamics of prosodic accommodation and lexical alignment as identical temporal spans are used to compute the levels of accommodation at these two linguistic levels.

2.2.2. Prosodic measurements

The model extracts a set of different acoustic parameters using Praat phonetic software and Matlab signal processing software. These parameters account for pitch range, articulation rate and voice intensity.

- Pitch range: median of fundamental frequency (med_f_0) and its standard deviation (sd_f_0). The med_f_0 and sd_f_0 are given on a logarithmic scale, the octave scale, $log2(Hertz)$, to facilitate gender comparisons. In order to avoid possible pitch tracking errors, pitch floor and pitch ceiling (when creating a Pitch Object) were set to the values $\mathrm{p}_{15} \cdot 0.83$ and $\mathrm{p}_{65} \cdot 1.92$ respectively, where p_{15} and p_{65} denote the 15th and 65th percentile respectively [28].
- Voice intensity: standard deviation of intensity (sd_Int) and its median (med_Int).
- Articulation rate : number of syllable nuclei per second ($syllsec$).

Speech/silent intervals and syllable nuclei are automatically annotated. Speech/silent intervals are detected using a method based on long-term modulation spectrum energy features [29]. Detection of syllable nuclei is performed using the method introduced in De Jong et al. [30], which is based on intensity peak detection of voiced segments of speech.

2.2.3. Quantification of prosodic accommodation dynamics

To measure prosodic accommodation for each parameter, we utilized Spearman's rank correlation coefficient, $\rho_{xy} \in [-1, 1]$. Large ρ values indicate strong accommodation over the analyzed fragment. The Spearman's coefficient analyses are executed on multiple levels of granularity. To investigate whether prosodic accommodation linearly increases over the course of the conversation, they are first computed for the first and second **H**alves of the conversations (***Experiment 1H***) as well as for the first, second and **T**hird parts of the conversations (***Experiment 1T***). To give a finer account of its dynamics, Spearman's coefficients are computed at **S**everal anchor points (***Experiment 1S***); 10 windows of the HYBRID feature extraction method, with a step size of 5, are grouped, which means that prosodic accommodation is calculated for a period of 110 seconds for every 50 seconds.

2.2.4. *Real vs. pseudo-interactions*

In order to investigate if the moments of prosodic accommodation are meaningful, we use a method similar to Ramseyer and Tschacher [31] and Ward & Litman [23]; our model creates a number of artificial conversations and from them computes Pseudo-accommodation coefficients (using same window sizes as described in 2.2.3). These coefficients are then compared, using t-tests, with those obtained from real conversations. If prosodic accommodation is significantly higher for Real than for Pseudo-interactions, this is taken as evidence that the prosodic accommodation captured is not random or accidental. The prosodic features as extracted by the model using the HYBRID method are used as input data for computing Pseudo-accommodation coefficients. Each set of prosodic features obtained for each speaker and each conversation is compared to each other set of prosodic features obtained for all other speakers and all other interactions. This allows us to compute the Pseudo-accommodation coefficients obtained for Pseudo-pairs for comparison with the Real accommodation coefficients from the real conversations.

2.3. Lexical alignment measurements

We adapt an extant method of assessing content synchrony in dialog [32, 33] to the needs of inspecting the dynamics of accommodation. The underlying method is applied to dialog transcripts which individuate discrete contributions (turns) of each participant into a linear sequence (an interrupted contribution becomes individual turns). The method quantifies each turn in sequence with respect to self-repetitions and allo-repetitions between it and preceding turns in the proportion of items shared vs. not shared. As a finite-register approach, a register is designated as a temporary store for each speaker's last dialog contribution and turns analyzed in sequence are compared with the contents of these registers, updated after processing of the turn. Turns prior to the immediately preceding contribution of each speaker are not considered (cf [22]). Transcripts are treated minimally: personal pronouns are normalized to a common form, but words are not lemmatized. Inspection of repetitions of $n-$grams includes $1 \leq n \leq 5$. The descriptive power of the method comes from comparing the actual repetition proportions in the dialog with those which emerge in randomized counterparts of the dialog in which turns are shuffled with respect to each other (10 shuffling for each analysis).

2.3.1. *Quantification of lexical accommodation dynamics*

Alignment can be quantified as the extent to which actual repetition exceeds expectations associated with random dialog based on the same contributions. While temporal overlap features have been explored using this method, it has not been deployed to examine the flow of content repetition over the course of dialogs; rather, the method has provided indices of synchronization in dialogs in their totality. Such analysis can be provided by processing transcripts using the method as is, but within a sequence of temporal windows over the dialog. That is, the analysis described above is conducted within each of the dialog windows. As it is performed for the prosodic level, lexical accommodation is first computed for the first and second **H**alves of the conversation (***Experiment 2H***); then, for the first, second and **T**hird parts (***Experiment 2T***); finally, at **S**everal anchor points (***Experiment 2S***), by dividing each conversation in windows of 110s with 50 second overlap. As in the HYBRID method, an utterance-sensitive model is used. At the lexical level, the temporal segmentation uses a turn-sensitive model, that respects the end of the turn, hence the propositional content of the interaction.

3. Results

3.1. Prosodic accommodation

Analyses ***across conversations*** reveal that on average, the levels of prosodic accommodation are low for the 5 conversations (M=0.07; SD=0,20) and that sd_f0 and med_f0 exhibit larger coefficients while sd_Int is smaller. The paired t-tests performed for ***Experiment 1H*** show no significant difference between the levels of prosodic accommodation in the first and second parts of the conversation, for all prosodic parameters ($p < .05$). Similarly, the paired t-tests performed for ***Experiment 1T*** reveal that prosodic accommodation levels are not significantly different in the first, second and third parts of the conversations, for all prosodic parameters ($p < .05$).

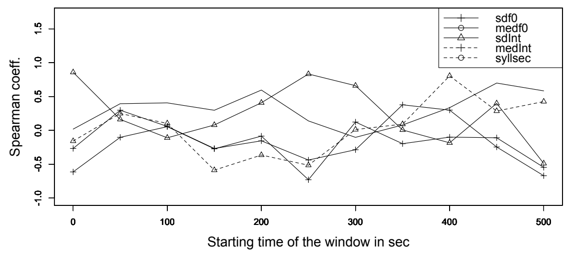

Figure 2: Prosodic accommodation levels obtained for each prosodic parameter, at several anchor points (every 50 sec, computed for a window of 110sec) for Conversation 12.

Experiment 1S reveals, for all conversations and all prosodic parameters, several phases of high and low prosodic accommodation levels (as exemplified in Fig. 2 for Conversation 12). In total (i.e. for all conversations), high accommodation levels[2] in terms of med_f0 represent 32% of the conversations; in terms of sd_f0 17%, in terms of med_Int 20%, in terms of sd_Int 9% and in terms of $syllsec$ 17%. The paired t-tests performed to compare Pseudo and Real Spearman's coefficients confirmed that the captured dynamics are not random, as Real coefficients are significantly larger than Pseudo coefficients ($t(687)$=2.79, $p < .01$).

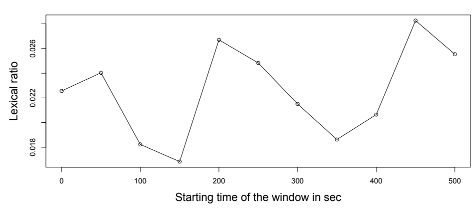

Figure 3: Lexical accommodation levels obtained at several anchor points (every 50 sec, window of 110sec) for Conv. 12.

[2]threshold determined using a local maxima based adaptive thresholding technique [34].

	C04						C12						C21						C32						C51					
Wn	L	SF	MF	MI	SI	S	L	SF	MF	MI	SI	S	L	SF	MF	MI	SI	S	L	SF	MF	MI	SI	S	L	SF	MF	MI	SI	S
1	+			+			+		+				+		+			+		+		+			+		+			
2														+	+											+	+			
3					+																									
4		+			+																+	+								
5							+	+													+									+
6							+		+										+									+		
7		+		+	+				+						+		+			+								+		
8					+	+												+		+					+		+			
9		+			+							+				+								+						
10		+			+			+						+							+	+				+				+

Table 1: Window-based comparison between lexical and prosodic accommodation. Wn stands for window number, L for Lexical accommodation; SF for accommodation in $sdf0$, MF in $medianf0$, MI in $medianInt$, SI in $sdInt$ and S in $syllsec$.

3.2. Lexical alignment

As for the prosodic level, analyses ***across conversations*** reveal that in average, the levels of lexical accommodation are low for the 5 conversations. Analyses performed for ***Experiment 2H*** and ***Experiment 2T*** do not reveal any increased lexical accommodation. ***Experiment 2S*** rather shows that lexical accommodation fluctuates over the course of the conversation (see Fig. 3). In total (i.e. for all conversations), lexical accommodation represents 16% of the conversations.

3.3. Temporal dynamics and agreement

Our results reveal that both prosodic and lexical accommodation do not linearly increase, but dynamically evolve over the course of a conversation. We therefore investigated whether these dynamics are aligned (***Experiment 3***) and whether they are correlated with inter-speaker agreement (***Experiment 4***).

Experiment 3. Regarding the alignment among prosodic parameters, ANOVA analyses reveal that accommodation fluctuations for the different prosodic parameters are correlated for sd_Int, med_Int and $syllsec$. The higher accommodation in terms of sd_Int, the higher in terms of med_Int and $syllsec$ ($sd\_Int \sim med\_Int$: F(1,67)=9.67; $sd\_Int \sim syllsec$: F(1,67)=9.06; $syllsec \sim med\_Int$: F(1,67)=4.06; $p < .05$). ANOVA analyses however show no correlation between the alignment measurements at the prosodic level and at lexical level: accommodation fluctuations at prosodic and lexical levels appear to be independent phenomena. Table 1 shows the overlap between lexical and prosodic accommodation. ***Experiment 4*** reveals no functional relation between agreement on an item and either lexical or prosodic accommodation.

4. Discussion

In this paper, we have presented a novel method for the automatic measurement of prosodic and lexical accommodation dynamics in social interaction. Using spoken dialogues of Scottish English, we have first investigated the increasing manifestation of accommodation by examining the amount of prosodic and lexical accommodation for halves and thirds of conversations. The results indicate that there is no significant difference between the first and second halves of all the conversations, nor between the first, second or third parts. We further analysed the evolution of prosodic and lexical accommodation at different anchor points and found that they vary several times over the course of all conversations. This corroborates previous findings reported for English [6, 20, 19] and Japanese [27] and supports the need for a model that can capture the dynamic manifestation of accommodation.

Results also reveal that pitch and intensity exhibit higher average number of accommodation phases than articulation rate, as previously observed in [20, 6, 27]. Pitch has been found to be strongly correlated with the activation dimension of emotional models [35, 36]. In [20, 6, 27], speakers' pitch accommodation is shown to be highly correlated to speakers' degree of involvement and affinity. It can be hypothesized that speakers show pitch accommodation much more than other types of prosodic adaptation as the human auditory system is very sensitive to changes in pitch. Speaker's states may be mainly expressed and recognized by changes in the amount of pitch accommodation. On the contrary, small changes in articulation rate may not be as well perceived, which would result in low levels of accommodation. Studies have shown that variations in speech rate are rather due to variations in the number of pauses and their mean duration than to variations in the actual articulation rate [37, 38]. While speakers' articulation rate is rather constant in nature, one may rather accommodate their speech in terms of pause duration, as was reported in [20, 6].

In this study, we have also investigated the temporal alignment of prosodic and lexical alignment. Our analyses reveal that these mechanisms are not aligned, suggesting they may serve different functions. It can be hypothesized that they evolve in a complementary manner. While lexicon accommodation would be augmented when new concepts are introduced, prosodic accommodation would serve the social function of communication and would be augmented at other instances of the conversation.

Finally, while previous studies have reported strong links between prosodic accommodation and speakers' socio-emotional states, this study does not reveal any correlation between prosodic / lexical accommodation and inter-speakers' agreement. This confirms previous findings by Vaughan et al [27]. Further work is however needed to confirm these preliminary results. Future work will aim to investigate accommodation and speakers' agreement at more localised levels, as at turn-takings and backchannels, which have been found to be important co-construction mechanisms in social interaction [5, 25].

5. Conclusions

In this paper, we propose a new methodology for analyzing the temporal evolution of accommodation in conversations. We have shown that accommodation fluctuates over the time of a conversation, supporting a model that can capture its dynamic manifestation. By comparing the dynamics of lexical and prosodic accommodation, we have also shown that their dynamics are not aligned. Our interpretation is that these mechanisms reflect different functions and therefore evolve in an independent manner.

6. Acknowledgements

This work was undertaken as part of the FASTNET project - Focus on Action in Social Talk: Network Enabling Technology funded by Science Foundation Ireland (SFI) 09/IN.1/2631.

7. References

[1] H. Giles, N. Coupland, and J. Coupland, "Accommodation theory: Communication, context, and consequence," *Contexts of accommodation: Developments in applied sociolinguistics*, pp. 1–68, 1991.

[2] J. Pardo, "On phonetic convergence during conversational interaction," *The Journal of the Acoustical Society of America*, vol. 119, p. 2382, 2006.

[3] G. W. Stanford and S. Webster, "A Nonverbal Signal in Voices of Interview Partners Effectively Predicts Communication Accommodation and Social Status Perceptions," *Journal of Personality and Social Psychology*, vol. 70, no. 6, pp. 1231–1240, 1996.

[4] J. Edlund, M. Heldner, and J. Hirschberg, "Pause and gap length in face-to-face interaction," in *Tenth Annual Conference of the International Speech Communication Association*, 2009, pp. 2779–2782.

[5] R. Levitan, A. Gravano, and J. Hirschberg, "Entrainment in speech preceding backchannels," *Proc. of ACL 2011*, 2011. [Online]. Available: http://www.aclweb.org/anthology/P/P11/P11-2020.pdf

[6] C. De Looze, C. Oertel, S. Rauzy, and N. Campbell, "Measuring dynamics of mimicry by means of prosodic cues in conversational speech," in *Proceedings of ICPhS*. Springer, 2011.

[7] S. Brennan, "Lexical entrainment in spontaneous dialog," *Proceedings of ISSD*, pp. 41–44, 1996.

[8] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue." *The Behavioral and brain sciences*, vol. 27, no. 2, pp. 169–90; discussion 190–226, Apr. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15595235

[9] A. Nenkova, A. Gravano, and J. Hirschberg, "High frequency word entrainment in spoken dialogue," in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*. Association for Computational Linguistics, 2008, pp. 169–172.

[10] H. Branigan, M. Pickering, J. Pearson, and J. McLean, "Linguistic alignment between people and computers," *Journal of Pragmatics*, vol. 42, no. 9, pp. 2355–2368, 2010.

[11] M. Pickering and V. Ferreira, "Structural priming: a critical review." *Psychological bulletin*, vol. 134, no. 3, p. 427, 2008.

[12] H. Giles, J. Coupland, and N. Coupland, *Contexts of Accommodation: Developments in Applied Sociolinguistics*, ser. Studies in Emotion and Social Interaction. Cambridge University Press, 1991. [Online]. Available: http://books.google.ie/books?id=vnKCc4QO7gQC

[13] P. Boylan, "Accommodation Theory Revisited," University of Rome III (Italy), Rome, Tech. Rep., 2004.

[14] F. Parrill and I. Kimbara, "Seeing and hearing double: The influence of mimicry in speech and gesture on observers," *Journal of Nonverbal Behavior*, vol. 30, no. 4, pp. 157–166, 2006.

[15] L. Tickle-Degnen and R. Rosenthal, "The Nature of Rapport and Its Nonverbal Correlates," *Psychological inquiry*, vol. 1, no. 4, pp. 285–293, 2007.

[16] J. L. Lakin and T. L. Chartrand, "Using nonconscious behavioral mimicry to create affiliation and rapport," *Journal of Psychological Science*, vol. 14, no. 4, pp. 334–339, 2003.

[17] C. Shepard, H. Giles, and B. Le Poire, "Communication accommodation theory," *The new handbook of language and social psychology*, no. 1.2, pp. 33–56, 2001.

[18] L. Miles, L. Nind, and C. Macrae, "The rhythm of rapport: Interpersonal synchrony and social perception," *Journal of experimental social psychology*, vol. 45, no. 3, pp. 585–589, 2009.

[19] B. Vaughan, "Prosodic synchrony in co-operative task-based dialogues: A measure of agreement and disagreement," in *Proceedings of Interspeech 2011*. ISCA, 2011, pp. 1865–1868.

[20] C. De Looze and S. Rauzy, "Measuring speakers' similarity in speech by means of prosodic cues: methods and potential," in *Proceedings of Interspeech 2011*. ISCA, 2011, pp. 1393–1396.

[21] K. W. Church, "Empirical estimates of adaptation: the chance of two noriegas is closer to p/2 than p 2," in *Proceedings of the 18th conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, 2000, pp. 180–186.

[22] D. Reitter and J. Moore, "Predicting success in dialogue," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Association for Computational Linguistics, 2007, pp. 808–815.

[23] A. Ward and D. Litman, "Automatically Measuring Lexical and Acoustic / Prosodic Convergence in Tutorial Dialog Corpora," in *ISCA Tutorial and Research Workshop*, 2007, p. 4.

[24] F. Bonin, N. Campbell, and C. Vogel, "Laughter and topic changes: Temporal distribution and information flow," in *In Proceedings of CogInfoCom 2012 3rd IEEE International Conference on Cognitive Infocommunications*, 2012, pp. 53–58.

[25] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[26] S. Kousidis, D. Dorran, C. McDonnell, and E. Coyle, "Times series analysis of acoustic feature convergence in human dialogues," in *Proceedings of Interspeech*, 2008.

[27] "Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction."

[28] C. De Looze, "Analyse et interprétation de l'empan temporel des variations prosodiques en français et en anglais contemporain," Ph.D. dissertation, Doctoral thesis, Université de Provence, 2010.

[29] H. Maganti, P. Motlicek, and D. Gatica-Perez, "Unsupervised speech/non-speech detection for automatic speech recognition in meeting rooms," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, vol. 4. IEEE, April 2007, pp. IV–1037–IV–1040.

[30] N. H. De Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior research methods*, vol. 41, no. 2, pp. 385–390, 2009.

[31] F. Ramseyer and W. Tschacher, "Nonverbal synchrony or random coincidence? how to tell the difference," *Development of Multimodal Interfaces: Active Listening and Synchrony*, pp. 182–196, 2010.

[32] C. Vogel and L. Behan, "Measuring synchrony in dialog transcripts," in *COST 2102 Training School*, 2011, pp. 73–88.

[33] C. Vogel, "Attribution of mutual understanding," *Journal of Law & Policy*, pp. 101–145, 2013.

[34] T. Giannakopoulos, A. Pikrakis, and S. Theodoridis, "A novel efficient approach for audio segmentation," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[35] P. N. Juslin and K. R. Scherer, *Vocal expression of affect*, ser. The New Handbook of Methods in Nonverbal Behavior Research, J. Harrigan, R. Rosenthal, and K. R. Scherer, Eds. Oxford, UK: Oxford University Press, 2005.

[36] R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression," *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 614–636, 1996.

[37] F. Goldman-Eisler, *Psycholinguistics: Experiments in spontaneous speech*. Academic Press, 1968.

[38] F. Grosjean and A. Deschamps, "Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation," *Phonetica*, vol. 31, no. 3-4, pp. 144–184, 1975.