Prosodic Synchrony in Co-operative Task-based Dialogues: A Measure of Agreement and Disagreement

Brian Vaughan[1]

[1]Speech Communication Laboratory,
Centre for Language and Communication Studies, Trinity College Dublin
bvaughan@tcd.ie

10.21437/Interspeech.2011-507

Abstract

Prosodic synchrony has been reported to be an important aspect of conversational dyads. In this paper, synchrony in four different dyads is examined. A Time Aligned Moving Average (TAMA) procedure is used to temporally align the prosodic measurements for the detection of synchrony in the dyads. An overlapping windowed correlation procedure is used to measure synchrony for six different prosodic parameters: mean pitch, pitch range, mean intensity, intensity range, centre of gravity and spectral slope. This study shows that a windowed correlation procedure better captures the dynamic nature of speech synchrony than a single measure across a whole conversation. This method also enables points of concurrent synchrony between prosodic parameters to be detected. Moreover, the synchrony of the prosodic parameters was considered in relation to levels of agreement and disagreement in the four dyads. Results show only one parameter in one dyad to be significantly correlated with agreement/disagreement.

Index Terms: synchrony, agreement and disagreement, conversational dyads

1. Introduction

Speakers engaged in conversation often adapt a similar speaking style in relation to prosodic, syntactic and lexical aspects [1], with research demonstrating speech similarity occurring in children as young as three [2]. Adapting ones speaking style to that of an interlocutor is an automatic unconscious process [3] and can be seen as an attempt to create a shared expressive space and a level of understanding [4]. While the exact functional role or roles of speech similarity has yet to be determined, it has been suggested that it is used to signal rapport and reduce social distance [5], to establish and maintain social relationships [6], and to ensure that a level of comprehension and understanding is achieved [7]. This is an important consideration in the further development of Human-Computer-Interfaces (HCI) and spoken dialogue systems [8] [9] [10].

Burgoon defines speech similarity as occurring when "the observed behaviors of two interactants, although dissimilar at the start of the interaction, are moving towards behavioral matching" [11] thus implying that it is a relatively linear phenomenon. Likewise, investigation and measurement of speaker similarity often assumes it to be a linear phenomenon that increases over the course of a conversation [6]. However, this approach and understanding of speech similarity fails to capture or understand its dynamic nature [12]. Furthermore, speech similarity is described using a myriad of terms, including accommodation[13], convergence [1], alignment [3] and coordination [14]. In this paper, we adopt the same definitions used by Edlund and Hirschberg [12]: similarity, convergence and synchrony. Similarity refers to the general phenomenon of speaker adaptation, convergence to the meeting of prosodic parameters at a shared point or points, and synchrony refers to the synchronous parallel movement of prosodic parameters between separate speakers. In order to capture the dynamics of speech synchrony, a moving correlation window was used. Moreover, this methodology was used to measure and detect points of concurrent synchrony across the different prosodic parameters.

As well as examining the form of speech synchrony, a preliminary examination of its functional nature in relation to agreement and disagreement is carried out. Some research has suggested that agreement can be detected using prosodic parameters (in conjunction with lexical and dialogue act labels) [15], as well as the level and type of agreement in a conversation [16]. It was hypothesised that, in a task-based cooperative situation, where agreement was necessary to advance, points of agreement in the conversation would (1) be correlated with points of synchrony for certain prosodic parameters and (2) that points of simultaneous synchrony across the parameters would correlate with points of agreement/ disagreement.

Prosodic synchrony is examined in four cooperative task-based dyads. Since speech similarity has been observed to be in dyads of same sex pairs [17], all four dyads examined herein were pairs of female friends. Six parameters were examined: mean pitch, pitch range, mean intensity, intensity range, spectral slope and centre of gravity (COG). COG and spectral slope are examined as measures of spectral energy distribution along with the four more frequently examined parameters. Synchrony in spectral energy is examined as it has been found to be related to important aspects of speech: it has been found to be positively related to the activation aspect of emotional states [18] and gives speech a 'harsh' and 'bright' voice quality [19].

2. Methodology

2.1. Data

The four conversations examined in this paper came from the Dublin Institute of Technology Emotional Speech Corpus (DITESC) [18]. Each dyad was recorded in a HD audio environment, with each interlocutor sitting in separate sound proof booths that contained a pair of headphones, a microphone and a computer screen. They were tasked with arranging 15 items on screen in order of importance for the survival of an imaginary shipwreck. There was a ten minute timer and a score, and participants were told that a point would be awarded for every item correctly ordered: unbeknownst to them, the score was being externally manipulated using a set scoring pattern. The record-

Copyright © 2011 ISCA

ings were in the audio modality only: participants could not see each other. Each dyad was approximately ten minutes long and both sides of the conversation were recorded as separate audio streams. No constraints were put on the participants speech and the dialogue was spontaneous. The recordings were converted from 192Khz 24bit to 16Khz 16bit for analysis.

2.2. Process

One of the problems with examining synchrony is the turn-based nature of conversational speech. This can be overcome using a Time Aligned Moving Average (TAMA) analysis window [10]. This method overcomes the turn-based problem by averaging values over a set of overlapping windows. As with [10], we used a 20 second window with a 10 second overlap between windows (Figure 1). This provides for a smoother set of prosodic contours over the duration of the conversation and ensures that the measurements are time aligned, thus aiding the detection of synchrony. Using this method, values were obtained for the six parameters discussed for each speaker in the four dyads using the Praat audio analysis software [20]. This resulted in values for the six parameters at approximately 65 points over the duration of the four conversations for each speaker. Each conversation was annotated for agreement using the agreement layer of the DAMSL annotation scheme . This layer details six agreement descriptors: maybe (m), hold (h), agree (a), agree-part (ap), reject (r) and reject-part (rp). In the DAMSL scheme, hold is defined as an utterance that seeks clarification (holding for more information) and maybe is defined as an utterance with no clear agreement or rejection [21]. For the purposes of this research, maybe and hold were amalgamated into a single neutral (n) descriptor, as neither signaled agreement or disagreement. As with the TAMA method, the annotation window was 20 seconds long with an overlap of 10 seconds to ensure an accurate determination of agreement in the dialogues and giving an agreement value for each of the 65 TAMA points. The agreement annotations were recoded into five numerical values to aid analysis: a=2, ap=1 h,m=0 rp=-1 and r=-2.

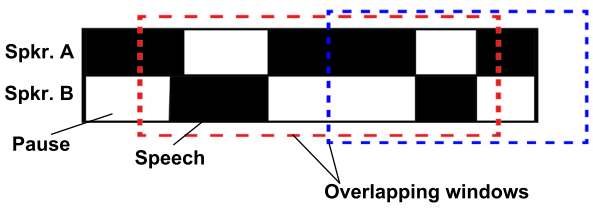

Figure 1: *Illustrative graphic of the TAMA method. Adapted from [10]*

2.3. Analysis

SPSS was used to carry out all statistical analysis [22]. Values for five of the six parameters were normalised using a z-means transformation: the pitch range values were calculated in semi-tones and so were not normalised in this way. For each dyad, a Pearson correlation coefficient was calculated for pairs of each one of the six prosodic parameter means for: a) the entire conversation and b) for a number of overlapping windows (essentially a windowed correlation procedure). Similar in nature to the TAMA windows, each contained ten points of values with a five point overlap. This resulted in approximately 12 overlapping correlation windows for each conversation, with each window corresponding to a 200 second period of conversation. Similarly, the overall level of agreement within each overlapping window was obtained by calculating the mode of the agreement scores within each individual window.

2.4. Results

2.4.1. Synchrony in dyads

For each dyad, a Pearsons correlation procedure was used to obtain correlation coefficients for all prosodic parameters, for the entire conversation (Table 1). Dyad one had significant correlation for average intensity and intensity range. Dyad two had a significant positive correlation for average F0 , F0 range , average intensity, intensity range, spectral slope and COG . Dyad three had a significant positive correlation for intensity range and COG. Finally, dyad four had significant correlation for average F0, intensity range and COG. Overall, intensity range was significantly correlated across all four dyads.

	Avg.F0	F0 Rng.	Avg.Int	Int.Rng	Sp.Slp	COG
D1	0.226	-0.074	**0.307***	**0.697****	0.139	0.07
D2	**0.711****	**0.433***	**0.795****	**0.752****	**0.323***	**0.340****
D3	0.191	0.142	0.239	**0.341****	0.17	**0.354****
D4	**0.377****	0.011	-0.083	**0.273***	0.102	**0.546****

Table 1: *Correlation coefficient values for all six parameters n each dyad. * signifies p<0.05 and ** signifies p<0.01*

2.5. Variations of synchrony

The windowed correlation procedure returned a number of significant correlations at various points across all parameters, for all dyads (* signifies p<0.05 and ** signifies p<0.01). Table 2 gives details of some of the most signifiant correlations for each of the six parameters in each dyad. Dyad one had a number of significant correlations for average F0, intensity range and COG. Dyad two had a number of significant correlations for F0, F0 range, average intensity and COG. Dyad three had a number of strong correlations for average intensity, intensity range and COG . Finally, dyad four had a number of strong correlations for average F0 and COG.

Dyad	F0	F0 Rng	Avg.Int	Int Rng	Sp.Slp	COG
1	0.618* 0.637* 0.65*		0.736*	0.884** 0.853** 0.871*	0.654*	0.605* 0.716* 0.775**
2	0.684* 0.772** 0.892**	0.765** 0.868** 0.831**	0.905** 0.891** 0.899**	0.433**	0.762**	0.627* 0.661*
3			0.642* 0.822** 0.883**	0.741* 0.717* 0.914**	0.764**	-0.685* -0.790** 0.743**
4	0.780** 0.736** 0.708*		-0.625*	0.638*		0.882** 0.925** 0.888**

Table 2: The most significant synchrony results for the windowed correlation procedure

2.6. Synchrony and agreement

The agreement scores for each of the 12 correlation windows were correlated with the synchrony values (the correlation coefficient values previously computed), for the six parameters, using a Pearson correlation procedure. The results given in Table 3 show that spectral slope in dyad three was significantly

negatively correlated with the agreement score. However, in general, there was little or no correlation for the six parameters in relation to agreement across all dyads.

Graphing the windowed correlations for each dyad with their points of agreement (Figures 2, 3, 4, 5) revealed a pattern of concurrent variation of the synchrony values, for some parameters, at certain points in the dyads. For example, the fifth correlation window in Figure 2 shows all parameters increasing their level of synchrony (an increased correlation coefficient value) at that point; in contrast, the fifth correlation window in Figure 4 shows a concurrent rise for only four of the six parameters at that point.

	Avg.F0	F0 Rng	Avg.Int	Int.Rng	Specl.Slp	COG
D1	0.03	-0.09	-0.27	-0.41	-0.34	-0.12
D2	0.12	-0.51	0.38	0.24	-0.22	0.15
D3	0.15	-0.28	0.57	0.44	**-0.67***	0.52
D4	0.17	0.06	-0.2	0.03	-0.11	0.17

Table 3: *Correlation coefficients for agreement scores and all parameters in all dyads. * signifies p<0.05.*

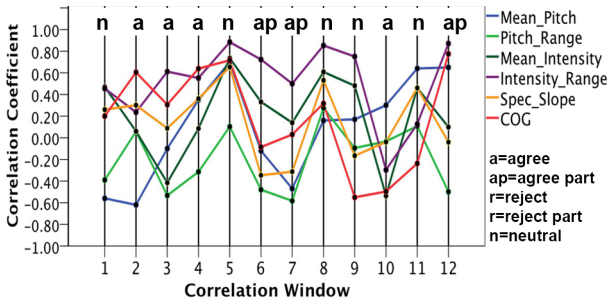

Figure 2: *Correlation contours for all six parameters in dyad 1 with agreement listed for each correlation window.*

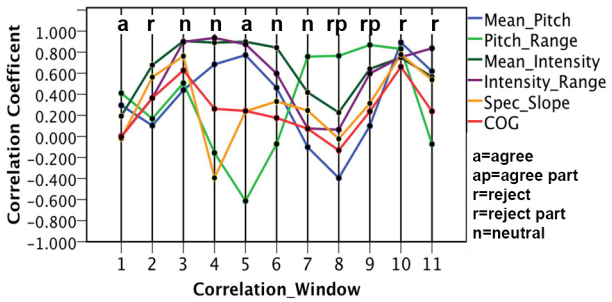

Figure 3: *Correlation contours for all six parameters in dyad 2 with agreement listed for each correlation window.*

To investigate whether this pattern of concurrent increases and decreases was a function of agreement, a mean synchrony score (the mean value of the correlation coefficient values) for the six parameters at each of the 12 points in the four dyads was calculated. Using Pearsons correlation coefficient, these mean values were correlated with the agreement mode calculated for each of the 12 points. There was no significant correlation for the mean concurrent synchrony values and the agreement values in the four dyads.

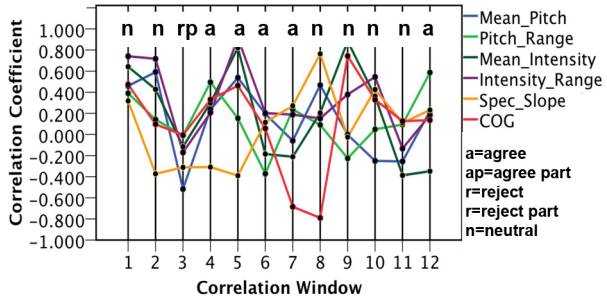

Figure 4: *Correlation contours for all six parameters in dyad 3 with agreement listed for each correlation window.*

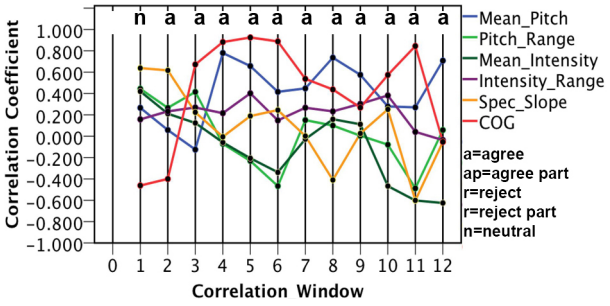

Figure 5: *Correlation contours for all six parameters in dyad 4 with agreement listed for each correlation window.*

3. Discussion

The results show that while a single correlation measurement across a whole conversation can sometimes reveal a general level of synchrony for prosodic parameters (Table 1), a moving overlapping correlation window can better capture the dynamic aspect of synchrony. In relation to agreement, the results show that there was no correlation for agreement and points of prosodic synchrony, except for spectral slope in dyad three. Moreover, points of concurrent synchrony across the prosodic parameters, for the four dyads, did not correlate with agreement.

The relationship between the interlocutors is most likely an important influence on the on the form and function of the conversation and the resulting prosodic synchrony and type of agreement. The participants in all four dyads were observed, and self-described, friends. The fact that no clear pattern of agreement and synchrony was found may be due to the interpersonal strategies utilised between friends rather than the adoption of a more general, global strategy.

While the conversations were spontaneous, the task based nature of the conversation may also have been a contributing factor: while there were points of disagreement in most of the conversations, it could be claimed that it is not a form of disagreement that would adversely affect the relationship of the interlocutors. Disagreement in a simple task-based game can not be considered the same as disagreement, for example, of the kind found between couples engaged in marriage counseling [16]. It might be more prudent to consider this type of disagreement as simply a lower level of agreement. While the patterns of concurrent variations in synchrony may be related to topic changes, the lengthy duration (200 seconds) of the correlation windows may preclude this from being a factor. However, it is prudent to consider that the observed patterns of synchrony between parameters may be due to phenomena not con-

sidered or quantified in this study. The dyads were taken from an emotional speech corpus and the tasks the interlocutors were engaged in were designed to elicit underlying emotional states [18]. This should be a consideration when interpreting the results and may also account for the lack of consensus across dyads in terms of the prosodic strategies employed to signal agreement/disagreement.

3.1. Future work

The windowed correlation procedure enabled concurrent points of increased and decreased synchrony between prosodic parameters to be detected. Three of the four dyads demonstrated a pattern of increased and decreased synchrony for the majority of the parameters investigated. More work is needed to understand this aspect in greater detail. It is quite possible that these patterns are manifestations of the interpersonal communicative strategies of the interlocutors. Different interlocutors may use different individual prosodic parameters, or a combination of parameters, to signal agreement, depending on their relationship, the social context, and the overall goal of the conversation. More work, using a larger data set, is also needed to conclusively demonstrate that a windowed correlation procedure is a better measure of dyadic synchrony than a single analysis of the whole utterance. Finally, the effect of the TAMA window length and the correlation window length needs to be investigated.

4. Conclusions

This study has shown that synchrony in spontaneous conversation is best understood as a dynamic phenomenon. The windowed correlation coefficient procedure outlined in this paper is a novel method for capturing this dynamic phenomenon. It has also shown that patterns of concurrent increased and decreased synchrony across different prosodic parameters can be detected using this method. Aside from a negative correlation for spectral slope in dyad three, no correlation was found between synchrony and agreement for individual prosodic parameters, nor for points of concurrent synchrony, in the four dyads. Finally, the results demonstrate that the spectral energy (spectral slope and COG) of interlocutors is also subject to dyadic synchrony .

5. 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/I2631. The author would like thank Dr. Celine de Looze and Professor Nick Campbell for comments and corrections.

6. References

[1] J. S. Pardo, "On phonetic convergence during conversational interaction," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, p. 2382, 2006. [Online]. Available: http://link.aip.org/link/JASMAN/v119/i4/p2382/s1&Agg=doi

[2] R. L. J. Street, N. J. Street, and A. . Van Kleek, "Speech convergence among talkative and reticent three year-olds," *Language Sciences*, vol. 5, no. 1, pp. 79–96, 1983.

[3] 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

[4] P. Boylan, "Accommodation Theory Revisited," Rome, pp. 1–26, 2004.

[5] C. A. Shepard, H. Giles, and B. A. Le Poire, *Communication accommodation theory*, 2nd ed., P. Robinson and H. Giles, Eds. Wiley, Jun. 2001.

[6] N. Suzuki and Y. Katagiri, "Prosodic alignment in human-computer interaction," *Connection Science*, vol. 19, no. 2, pp. 131–141, Jun. 2007.

[7] S. Garrod and M. J. Pickering, "Why is conversation so easy?" *Trends in Cognitive Sciences*, vol. 8, no. 1, pp. 8–11, Jan. 2004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S136466130300295X

[8] L. Bell, J. Gustafson, and M. Heldner, "Prosodic adaptation in human-computer interaction," in *Proceedings of ICPhS*, vol. 3. Citeseer, 2003, pp. 833–836.

[9] R. Coulston, S. Oviatt, and C. Darves, "Amplitude convergence in childrens conversational speech with animated personas," in *Proceedings of the 7th International Conference on Spoken Language Processing*, vol. 4. Citeseer, 2002, pp. 2689–2692.

[10] S. Kousidis, D. Dorran, C. Mcdonnell, and E. Coyle, "Convergence in Human DialoguesTime Series Analysis of Acoustic Feature," in *Proceedings of SPECOM 2009*, St. Petersburg, Russia., 2009, p. 2.

[11] J. K. Burgoon, L. A. Stern, and L. Dillman, *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge University Press, 1995, no. Cambridge, UK. [Online]. Available: http://www.loc.gov/catdir/description/cam026/94043249.html

[12] J. Edlund, M. Heldner, and J. Hirschberg, "Pause and gap length in face-to-face interaction," in *Proceedings of Interspeech 2009,*, 2009, pp. 2779–2782.

[13] H. Giles, N. Coupland, and J. Coupland, *Accommodation theory: Communication, context, and consequence*. Cambridge University Press, 1992, pp. 1–68.

[14] 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, Dec. 2002. [Online]. Available: http://jls.sagepub.com/cgi/doi/10.1177/026192702237953

[15] S. Germesin and T. Wilson, "Agreement detection in multiparty conversation," in *Proceedings of the 2009 international conference on Multimodal interfaces*, no. Id. New York, New York, USA: ACM, 2009, pp. 7–14. [Online]. Available: http://portal.acm.org/citation.cfm?id=1647319

[16] M. Black, A. Katsamanis, C. Lee, A. Lammert, B. Baucom, A. Christensen, P. Georgiou, and S. Narayanan, "Automatic Classification of Married Couples' Behavior Using Audio Features," in *Eleventh Annual Conference of the International Speech Communication Association*, Makuhari, Chiba, Japan, 2010, pp. 3–6. [Online]. Available: http://www.isca-speech.org/archive/interspeech_2010/i10_2030.html

[17] G. Bailly and A. Lelong, "Speech dominoes and phonetic convergence," in *INTERSPEECH-2010*, no. September, 2010, pp. 1153–1156.

[18] B. Vaughan, "Naturalistic Emotional Speech Corpora with Large Scale Emotional Dimension Ratings." Ph.D. dissertation, Dublin Institute of Technology, 2011.

[19] C. L. Breazeal, *Designing Sociable Robots (Intelligent Robotics & Autonomous Agents)*. MIT Press, 2002. [Online]. Available: http://www.amazon.co.uk/Designing-Sociable-Intelligent-Robotics-Autonomous/dp/0262025108

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2006.

[21] M. G. Core, , and J. F. Allen, "Coding Dialogs with the DAMSL Annotation Scheme." in *In Working Notes of AAAI Fall Symposium on Communicative Action in Humans and Machines.*, Boston, MA., 1997.

[22] IBM, "Home page for PASW/SPSS software," Chicago, 2010. [Online]. Available: http://www.spss.com/software/statistics/