

# Prosodic Convergence, Divergence, and Feedback: Coherence and Meaning In Conversation

Li-chiung Yang

Faculty of Humanities  
Tunghai University, Taichung, Taiwan  
yang\_lc@thu.edu.tw,  
lichiong.yang@gmail.com

## Abstract

A key goal for participants in language communication is to bring about a mutually shared experience of ideas, event narratives, and emotional responses. This goal is achieved not only through the exchange of lexical meaning, but also through interactive signaling to coordinate information status. Our results show that prosodic synchrony (convergence) and dissynchrony (divergence) both occur in conversation, and that synchrony is achieved gradually as participants cooperate to build up a shared information and involvement state. Our analysis further indicates that feedback is a critical component of cooperative adaptation to new information, bringing about convergent speaker states.

## 1 Introduction

Human language provides an especially cogent platform for studying the phenomenon of imitative and convergent behaviors in human communication, as speech communication integrates a complex mix of cognitive, emotional, and interactive social processes that are expressed in a number of different forms: the language specific choice of lexical items to communicate meaning, visually-based information exchange of gestures and facial expressions, and the shaping of the oral and aural environment through variations in prosodic flow. Scientific studies have shown convergent behavior in body movements and gesturing in conversation (Condon and Sander, 1974; Nagaoka, et al., 2007, Campbell and Stefan, 2010), and in speech (Gratch, et al., 2007; Jonsdottir, et al., 2007, Buschmeier, et al. 2011, Lelong and Bailly, 2011; Heylen, et al., 2011; Ward, 2006), and focused on their role in

creating harmony and rapport between conversational participants through the use of feedback markers, and through timing and frequency of non-verbal facial and movement gesturing (Lelong and Bailly, 2011; Heylen, et al., 2011).

Spontaneous conversation is multi-functional in both its goals and processes: the most evident goal of transmitting information simultaneously carries a social goal of building rapport and the sharing of attitudes and emotions towards the information transmitted. In the conversational process, speakers provide propositional and emotional and information through prosody, gesturing, and feedback, and engage in interactional probing to build a shared knowledge state and guide topic in a mutually desired direction. Prosody plays a key role in this process, as it provides a powerful and informative resource to communicate multiple levels of coherence and meaning by providing a direct and immediate link to fundamental expressive states.

## 2 Goal and perspective

The current study presents our results on prosodic convergence and divergence in spoken dialogues, drawing from extended conversational data in Mandarin Chinese. Because of the multi-dimensional goals at work in language, synchrony is approached as both building social interactional harmony, and also reflecting informational, organizational and expressive processes in conversations. The coherence achieved in a successful dialog is a shared coherence, one that is constructed through interactions of participants to discover and overcome respective inadequacies of information state. Thus, in addition to imitative speech patterns, prosodic convergence and divergence are considered as information-rich

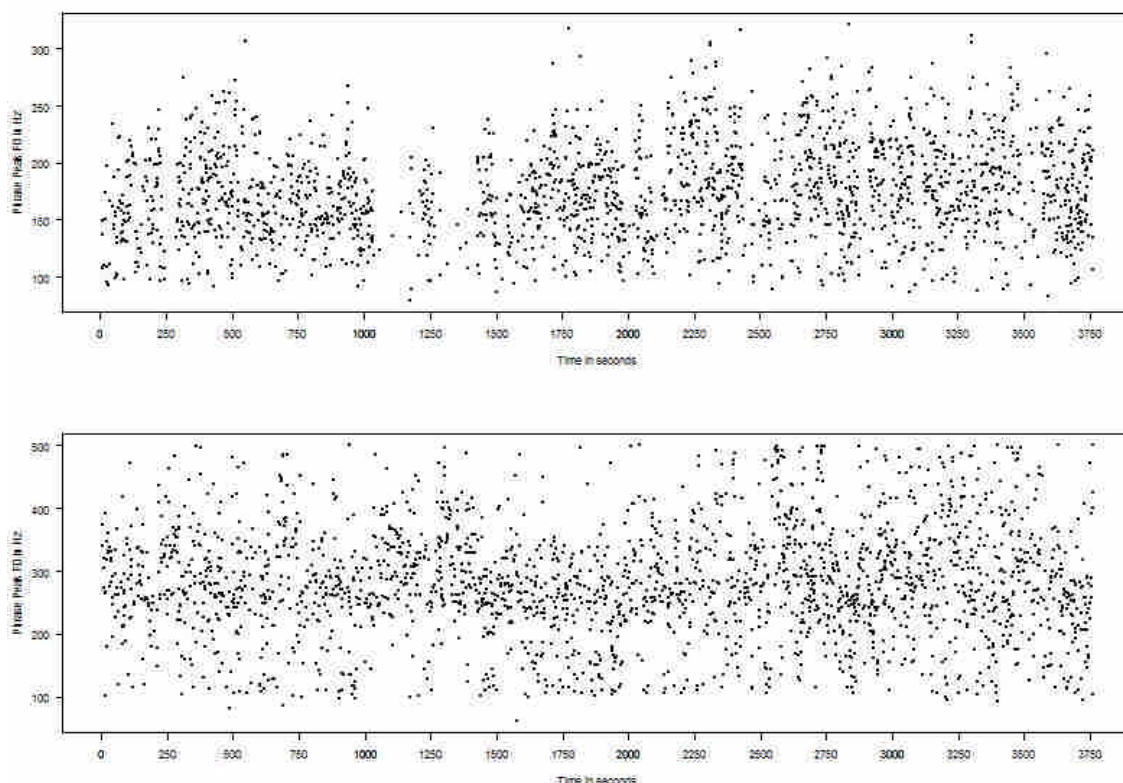


Figure 1: Times series of phrase peak  $f_0$  in Hz for the male speaker (top), and female speaker (bottom) in a 62-minute long conversation. The x-axis is time in seconds and the y-axis is pitch in Hz.

patterns that speakers use to monitor comprehension, communicate disinterest or encouragement, and signal different levels of agreement and judgment on topic.

### 3 Data and methodology

For this study, our data corpora consist of two extended spontaneous conversations in Mandarin Chinese, each approximately one hour in length. The Mandarin data are a subset of Academia Sinica’s Mandarin Conversational Dialogue Corpus (MCDC)<sup>1</sup> of natural conversations between newly-met participants (Tseng, 2004). The conversations were recorded in stereo in a quiet room and both conversations selected were mixed-gender pairs, with 1 male and 1 female participant. For these conversations, there were no preset topics and the speakers were free to talk about anything that arises naturally from the communication process (see Tseng, 2004 for a detailed account of the recording

and processing procedures). For ease of processing, each conversation was subdivided into 20 subsections, with approximately 3 minutes per episode. For the current study, the conversational data were further segmented to the phrase level, i.e. phrase-size chunklets, based on a combination of lexical, syntactical, semantic as well as acoustical criteria, and target tokens of interest were annotated and extracted. Measurements of fundamental frequency ( $f_0$ ) and amplitude were automatically computed, and normalized to each speaker’s pitch mean and range. For each speaker and each phrase, low, average, and high values for both  $f_0$  and amplitude were extracted as a means to show global pitch and amplitude movement variation. The acoustic measurements were then examined and correlated with incidence of feedback response and speaker interactions. Altogether there were 1,246 phrases for the female speaker and 2,273 phrases for the male speaker in mcdc01, and 2,256 for the female speaker and 2,014 for the male speaker in mcdc05, resulting in 3,519 phrases for mcdc01 and 4,280 for mcdc05 with a total of 7,799 phrases. Figure 1 shows the Times series of phrase peak  $f_0$  in Hz for the both speakers.

<sup>1</sup> For detailed information about the Mandarin Conversational Dialogue Corpus (MCDC), please see [http://mmc.sinica.edu.tw/mcdc\\_e.htm](http://mmc.sinica.edu.tw/mcdc_e.htm)

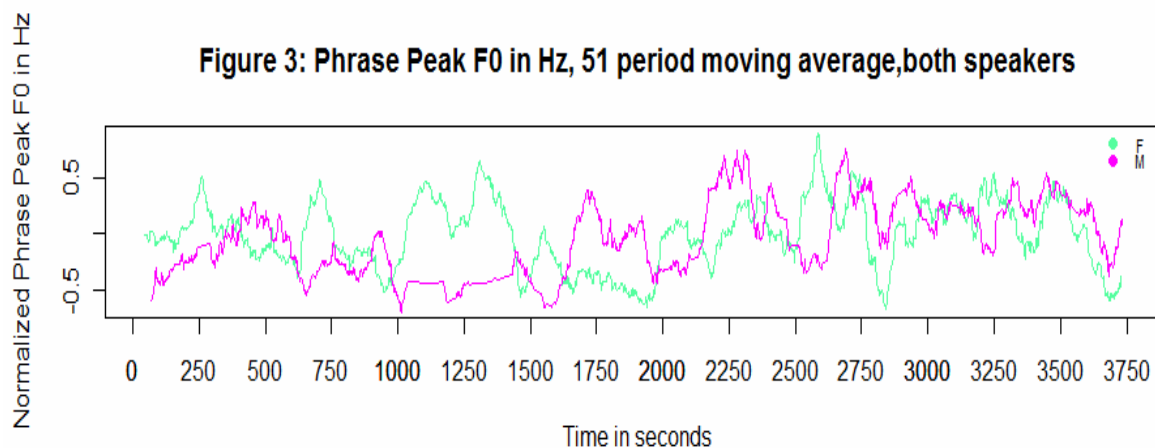


Figure 2: Normalized phrase peak with 51 period moving average for both speakers in MCDC conversation

## 4 Analysis and results

### 4.1 Conversational structure and prosodic convergence

By prosodic convergence (synchrony), we mean that speakers often use corresponding or matching movements in phrase pitch level to signal agreement on the current topic hierarchy. Our results show that both convergence and divergence in prosody occur at both local inter-phrase level pitch level changes, as well as over dialogue sections extending globally across topics and subtopics. Figure 2 compares the normalized global phrase movement for the two speakers in the Mandarin dialogue depicted above, and the moving correlation at lag 0.

The main pattern for the Mandarin conversational corpora is that prosodic synchrony is arrived at gradually, with an initial probing stage where topic is negotiated, followed by mixed convergence and divergence as options are explored or overturned from a one-sided viewpoint, until speakers arrive at a mutually fulfilling topic theme, where convergence is frequent. Near conversation end, participants converge in a descending pitch pattern in a shared recognition of the coming conclusion.

By comparison to talks between friends, conversations between newly-met participants may be more susceptible to lags in convergence, as speakers work to construct a common conversational outlook. The current results indicate that prosodic lags go in both directions, as speaker roles change and new topics are brought up during the course of the conversation. At the local level, prosodic synchrony at phrase-to-phrase pitch

movement is common: convergence is associated with agreement or encouragement of topic, whereas divergence is associated with disagreement, doubt, or non-interest.

### 4.2 Feedback and prosodic convergence

Speaker role was found to be important in the incidence and location of feedback tokens with respect to the prosodic patterns. Feedback markers of high interest or surprise such as 'oh', and encouraging markers such as 'um' or 'umhum' occur more frequently in areas of high pitch and convergence, and less frequently in divergent prosodic sections. The marker 'dui' right occurs more frequently in areas of convergence and stretches of extended rise as the hearer provides added encouragement or confirmation respectively. Thus, feedback markers often provide explicit marking of the same underlying relational states that are provided by synchrony phenomena.

Our data indicate that speakers differ greatly in their use of feedback in different conversations, in both frequency and distribution, as depicted in Figures 2-3, which show the incidence of feedback for *oh* and *dui* across 20 continuous 3-minute episodes of conversations mcdc05 (due to space limitation, the 01 figures are not shown in this paper). As seen in Tables 1-2, the frequency of feedback for conversation mcdc01 is less than half of that for conversation mcdc05: for mcdc05, the male and female *oh* and *dui* feedback over all episodes totals 305, or 43.1% of mcdc05's total feedback frequency of 707. The use of feedback by

the female speaker in mcdc01 is especially low for both *oh* and *dui*, with only 111 instances total.

Comparing the speaker usage of feedback markers *dui* and *oh*, we can see that in these two conversations, both female speakers' use of *oh* is relatively greater than *dui*: there are 82 *oh*'s vs. 29 *dui*'s in mcdc01, and 207 *oh*'s vs. 169 *dui*'s in mcdc05. Conversely, the male speakers' use of *dui* in both conversations is *much* greater than their use of *oh*: 171 *dui*'s vs. 23 *oh*'s in mcdc01 and 322 *dui*'s vs. 9 *oh*'s in mcdc05. The males have a striking near exclusion of the use of *oh* in both conversations.

MCDC01	Female		Male	
Token	<i>oh</i>	<i>dui</i>	<i>oh</i>	<i>dui</i>
Counts	82	29	23	171
Subtotal	111		194	
MCDC05	Female		Male	
Token	<i>oh</i>	<i>dui</i>	<i>oh</i>	<i>dui</i>
Counts	207	169	9	322
Subtotal	376		331	

Table 1: Counts of feedback markers *oh* and *dui* by speaker for 2 conversations, mcdc01 and mcdc05

MCDC01	Female		Male	
Token	<i>oh</i>	<i>dui</i>	<i>oh</i>	<i>dui</i>
Mean	4.05	1.45	1.15	8.55
Stdev	3.99	1.86	1.31	3.67
MCDC05	Female		Male	
Token	<i>oh</i>	<i>dui</i>	<i>oh</i>	<i>dui</i>
Mean	10.35	8.45	0.45	16.1
Stdev	7.84	3.32	0.80	11.11

Table 2: Mean counts of feedback markers *oh* and *dui* by speaker for 2 conversations, mcdc01 and mcdc05

The temporal distribution of feedback within conversations also varies greatly by speaker and conversation. In mcdc01, the female speaker has a higher concentration of feedback responses in the first 7 episodes of the conversation, and gradually reduces her feedback in the latter half of the conversation, while the male's feedback distribution is more uniform across the conversation. By contrast, the feedback for both speakers in mcdc05 occurs with higher frequency across the conversation, and also exhibits cyclical behavior.

### 4.3 Patterns of feedback distribution in conversation

Feedback markers are key interactive signals that communicate the adequacy of information exchange, and the distribution of specific markers is closely linked to their specific functions and to the

emotional and involvement states of speakers and interactivity level of the conversation. Imbalances in participant state commonly give rise to different degrees of cognitive certainty or uncertainty, and feedback provides immediate signals to speakers that adjustment or restatement of information may be necessary.

For example, *oh*, *dui*, and *umhum* are three of the most frequent feedback markers in Mandarin (Tseng, 2004; Yang, 2006), and each signals different degrees of cognitive uncertainty and receptivity towards communicated information. While *oh* functions as a response to information, marking surprise, unexpectedness or newness, and necessitating a cognitive adjustment (reorientation), *dui* acts as confirmation and agreement to information received, and implied as already known or accepted. The predominant function of *umhum* (*uhhuh*), on the other hand, is expression of acknowledgment or encouragement. Thus, *oh*, *dui*, and *umhum* each has its unique different functions and occurs under different informational environments. The specific functions of these markers have great significance for their frequency and distribution in any given conversation.

If we take a closer look at the following figures and tables where we tabulated and plotted the occurrences of these three feedback markers through time in mcdc05 by speaker, we can see a clearer pattern: there is a clear gender difference and preference for the use of these markers. In this conversation, the male speaker has a mean of 7.6 instances of *umhum* per episode, just 26% of the female's mean frequency of 29.1 for the same marker. A similar large gap exists for *oh*, with the female speaker having about 20 times as many *oh*'s as the male speaker. Conversely, the male speaker used *dui* about twice as often as the female speaker, with mean episode counts of 16.1 for the male speaker vs. 8.45 for the female.

The relationships for *oh* and *dui* in mcdc05 are consistent with the results for mcdc01, with female *oh* and male *dui* having the higher relative frequencies, as presented earlier. The much greater use of *oh* and *umhum* by the female and greater use of *dui* by the male speaker presented here suggest that there exists some social-cultural expectations of greater male control and greater female supportiveness in male-female social interactions, and this feature might be especially marked in conversations where politeness and role-conformity could be expected to exert greater force.

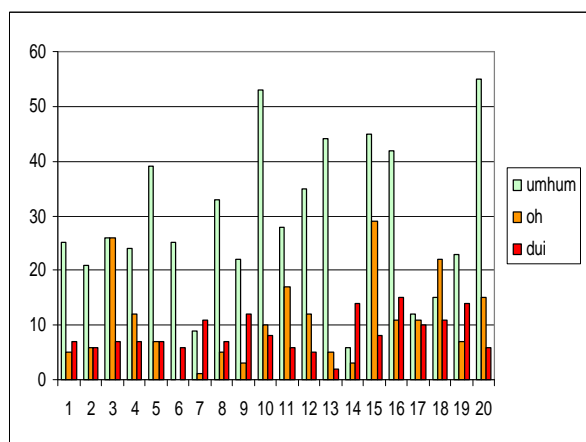


Figure 3: Frequency counts of feedback markers *umhum*, *oh*, and *dui* of the female speaker in mcdc05 by episode

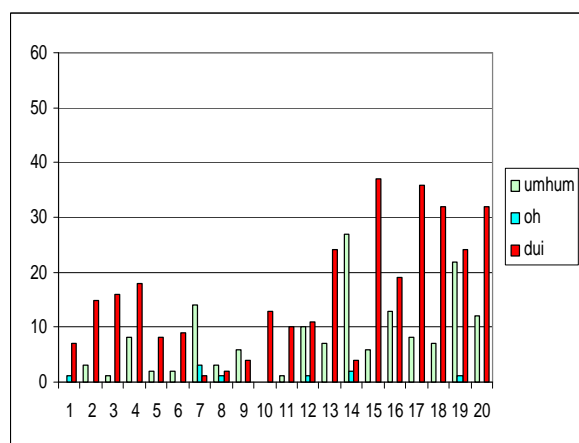


Figure 4: Frequency counts of feedback markers *umhum*, *oh*, and *dui* of the male speaker in mcdc05 by episode

Feedback	Male			Female		
	umhum	oh	dui	umhum	oh	dui
Mean	7.6	0.45	16.1	29.1	10.35	8.45
Stdev	7.05	0.80	11.11	13.57	7.84	3.32
Counts	152	9	322	582	207	169

Table 3: Mean counts of feedback markers *umhum*, *oh* and *dui* by speaker for mcdc05

Gender	mcdc01				mcdc05			
	Female		Male		Female		Male	
Token	oh	dui	oh	dui	oh	dui	oh	dui
Counts	82	29	23	171	207	169	9	322
Subtotal	111		194		376		331	
Total	305				707			

Table 4: Counts of feedback markers *umhum*, *oh* and *dui* by speaker for mcdc01 and mcdc05

#### 4.4 Feedback and changing speaker state

The progression of feedback use in this conversation follows the activities of each speaker as they interact to bring about a successful conversation. At the start of this conversation, both participants explore several topics in sequence, with the female speaker more open in sharing information and responsive to the male speaker in the first half of the conversation. The initial topics serve as self-introductions and as search activity to arrive at a mutually satisfying topic.

As shown in Figures 3-4 and Table 5, the relatively low feedback activity of the male speaker in the initial episodes is matched by a high frequency of encouraging *umhum* feedback by the female speaker, to provide support for the male

speaker. In episodes 7-9, the female speaker started an extended narration to tie her own experience and viewpoint to the male speaker’s account, which is an essential cohesion-building strategy. The male speaker stays relatively silent during this period, and this results in very low use of both *oh* and *dui* for both speakers in those episodes. After the repeated rapport-building activities by the female speaker, the male speaker gradually becomes more open and emotionally expressive, and this transformation is reflected in his increased use of feedback from that point on.

In discourse, participants may unintentionally hit an area of high interest, and participants may become very involved. In this conversation the topic hits a major turning point in episode 14. At that point, both participants suddenly discover something unexpected but highly relevant and meaningful to both of them, and this transforms the nature of the conversation, with a high intensity of involvement by both participants, as evidenced by the greatly increased use of feedback markers from that episode on. This effect is especially dramatic for the male speaker.

As can be seen in Table 6, the male speaker’s use of feedback increases greatly after the turning point, with *umhum* increasing to over 3 times its pre-turning point average, and *dui* about 2.5 times, while his use of *oh* decreased slightly, indicating his increased confidence and certainty associated with this newfound identity. By comparison, the female speaker’s use of feedback is also increasing, and consistent over the conversation, with *umhum* occurring at the same frequency, and *oh* and *dui* increasing by about 2/3.



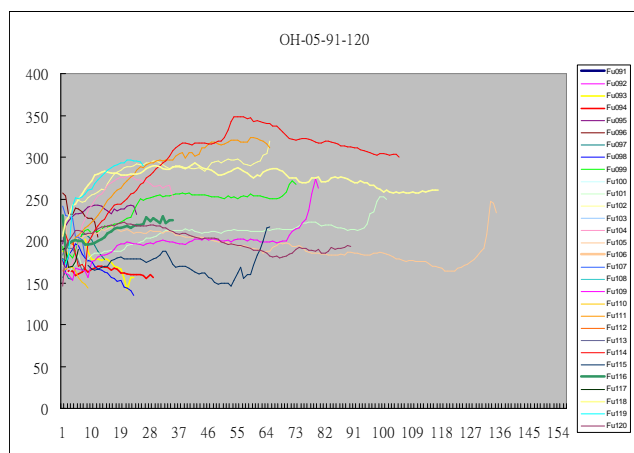


Figure 5: A selection of 30 instances of the female speaker’s *ohs* in mcdc05, approximately occurring from episode 11 to the beginning of episode 14, showing different forms relating to different states and functions in spontaneous conversation

The result of our finding suggests that the use of feedback to encourage rapport can consist of a feedback loop that increases rapport between participants and leads to more synchronous feedback patterns over time. Our finding further provides evidence that convergent patterns occur as speakers cooperatively achieve a shared common ground, and that feedback is a key element of how speakers reach this goal in communication (Yang, 2006).

Our findings show that speaker role plays a significant role in the incidence and location of feedback markers with respect to the prosodic patterns. Feedback markers of high interest or surprise such as *oh*, and encouraging markers such as *um* or *umhum* occur more frequently in areas of high pitch and convergence, and less frequently in divergent prosodic sections. The marker *dui* ‘right’ occurs more frequently in areas of convergence and stretches of extended rise as the hearer provides added encouragement or confirmation respectively. Thus, feedback markers often provide explicit marking of the same underlying relational states that are provided by synchrony phenomena. Figure 5 provides an illustration of how meaning is effectively encoded in such short feedback utterances for spoken communication.

Ep.	Male			Female		
	uhm	oh	dui	uhm	oh	dui
1	0	1	7	25	5	7
2	3	0	15	21	6	6
3	1	0	16	26	26	7
4	8	0	18	24	12	7
5	2	0	8	39	7	7
6	2	0	9	25	0	6
7	14	3	1	9	1	11
8	3	1	2	33	5	7
9	6	0	4	22	3	12
10	0	0	13	53	10	8
11	1	0	10	28	17	6
12	10	1	11	35	12	5
13	7	0	24	44	5	2
14	27	2	4	6	3	14
15	6	0	37	45	29	8
16	13	0	19	42	11	15
17	8	0	36	12	11	10
18	7	0	32	15	22	11
19	22	1	24	23	7	14
20	12	0	32	55	15	6
T			483			958

Table 5: Counts of feedback markers *umhum*, *oh* and *dui* by speaker and episode: mcdc05

	Male			Female		
	umhum	oh	dui	umhum	oh	dui
Pre TP	57.0	6.0	138.0	384.0	109.0	91.0
Post TP	95.0	3.0	184.0	198.0	98.0	78.0
Pre mean	4.4	0.5	10.6	29.5	8.4	7.0
Post mean	13.6	0.4	26.3	28.3	14.0	11.1

Table 6: Mean counts of feedback markers *umhum*, *oh* and *dui* by speaker, before and after the turning point for mcdc05

## 5 Conclusions

Our analysis suggests that prosodic synchrony phenomena occur as a mirror of topically and emotionally synchronized or dis-synchronized participant states and that convergence and divergence phenomena are not only strategies to encourage rapport, but also act as organizational indicators providing key information on the degree of understanding, on emotional synchrony, and on the perceived status of a mutually fulfilling topic flow. This universal feature is essential to communication and interaction, and should be utilized greatly in current multimodal communication environment research.

## Acknowledgment

The research is supported by National Science Council of Taiwan under grant no. NSC97-2410-H-029-026. The funding support is gratefully acknowledged by the author.

## References

- Amélie Lelong & Bailly, Gérard. 2011. Study of the phenomenon of phonetic convergence thanks to speech dominoes. In A. Vinciarelli, K. Vicsi, C. Pelachaud and A. Nijholt (eds.) *Analysis of verbal and nonverbal communication and enactment: the processing issue*, 280–293.
- Buschmeier, H., Z. Malisz, M. Włodarczak, S. Kopp, and P. Wagner. 2011. 'Are you sure you're paying attention?'–'Uh-huh'. Communicating understanding as a marker of attentiveness. *Proceedings of INTERSPEECH 2011*, Florence, Italy, 2057–2060.
- Campbell, Nick, and Scherer, Stefan. 2010. Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. *Proceedings of Interspeech 2010*, 2546–2549.
- Chika Nagaoka, Masashi Komori, & Sakiko Yoshikawa. 2007. Embodied synchrony in conversation. In Toyoaki Nishida (ed.) *Conversational informatics: an engineering approach*, Wiley Series in Agent Technology. John Wiley & Sons. 331–352.
- Condon, W. S., Sander, L. W. 1974. Neonate movement is synchronized with adult speech, interactional participation and language acquisition. *Science*, Vol.183, 99–101.
- D. Heylen, Bevacqua, E., Pelachaud, C., Isabella Poggi, Gratch, J., and Schröder, M. 2011. Generating listening behaviour. In P. Petta et al. (eds.), *Emotion-Oriented Systems*, Cognitive Technologies, Springer-Verlag Berlin Heidelberg. 321–347
- Emanuel Schegloff. 1982. Discourse as an interactional achievement: some uses of *uh huh* and other things that come between sentences. *GURT 1981 Analyzing Discourse: Text and Talk*, ed. by D. Tannen. Georgetown University Press, 71–93.
- G. Jonsdottir, Gratch, J., Fast, E., and Thórisson, K. 2007. Fluid semantic back-channel feedback in dialogue: challenges and progress. In C. Pelachaud et al. (Eds.): *IVA 2007*, LNAI 4722, Springer-Verlag Berlin, Heidelberg, 154–160
- Herbert Clark. 1996. *Using language*. Cambridge: Cambridge University Press.
- Howard Giles and P. Smith. 1979. Accommodation theory: Optimal levels of convergence. In H. Giles & R. St. Clair (Eds.), *Language and social psychology*, 45-65. Oxford: Blackwell.
- Jennifer Pardo. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustic Society of America* 119(4): 2382–93.
- Jennifer Pardo. 2010. Expressing oneself in conversational interaction. In *Expressing oneself/Expressing one's self: communication, cognition, and identity*, ed. E. Morsella (Taylor & Francis London), 183–196.
- Kawai Chui. 1994. *Information Flow in Mandarin Chinese Discourse*. PhD dissertation. National Taiwan Normal University.
- Li-chiung Yang. 1995. *Intonational Structures of Mandarin Discourse*, PhD dissertation, Georgetown University.
- Li-chiung Yang. 2006. Integrating prosodic and contextual cues in the interpretation of discourse markers. In *Approaches to Discourse Particles*, ed. By Fischer, Kerstin, Elsevier, 265–297.
- Lixing Huang, Morency, L., and Gratch, J. 2010. Learning backchannel prediction model from parasocial consensus sampling: A Subjective Evaluation. *The 10th International Conference on Intelligent Virtual Agents (IVA 2010)*, Philadelphia, PA, USA, 159–172.
- Lixing Huang and Gratch, J. 2012. Crowdsourcing backchannel feedback: understanding the Individual Variability from the Crowds. *The Interdisciplinary Workshop on Feedback Behaviors in Dialog*, Portland, Oregon.
- Nigel Ward. and Tsukahara, W. 2000. Prosodic Features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 23, 1177–1207.
- P. Sadler, Ethier, N., Gunn, G.R, Duong, D, and Woody, E. 2009. Are we on the same wavelength? Interpersonal complementarity as shared cyclical patterns during interactions. *Journal of Personality & Social Psychology*, 97(6):1005–20.
- Jens Allwood, Nivre, J. and Ahlsén, E. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26.
- Jonathan Gratch, Wang, N., Gerten, J., Fast, E., & Duffy, R. 2007. Creating rapport with virtual agents. *The 7th International Conference on Intelligent Virtual Agents*. Paris, France, 125–138.
- S. P. Gill. 2012. Rhythmic synchrony and mediated interaction: towards a framework of rhythm in embodied interaction. *AI & Society*, 27, 111–127.
- Shu-Chuan Tseng. 2003. Taxonomy of spontaneous speech phenomena in Mandarin conversation. In *SSPR-2003*, ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition, Tokyo Institute of Technology, Tokyo, Japan.
- Shu-Chuan Tseng. 2004. Processing spoken Mandarin corpora. *Traitement automatique des langues. Special Issue: Spoken Corpus Processing*, 45(2): 89–108.
- Shu-Chuan Tseng. 2008. Spoken corpora and analysis of natural speech. *Taiwan Journal of Linguistics*, vol. 6.2, 1–26.
- Susan Brennan, & Herbert Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493.
- Wallace Chafe. 1994. *Discourse, consciousness and time*. Chicago: University of Chicago Press.
- Wallace Chafe. 2008. The analysis of discourse flow. In *The Handbook of Discourse Analysis*. Oxford: Blackwell, 673-687.