Prosodic convergence in Italian game dialogues

Michelina Savino[1], Loredana Lapertosa[1], Alessandro Caffò[1], Mario Refice[2]
[1]Dept. of Education, Psychology, Communication, University of Bari, Italy
[2]Dept. of Electrical and Information Engineering, Polytechnical Univ. of Bari, Italy
https://doi.org/10.36505/ExLing-2016/07/0034/000293

Abstract
In this study we explore the manifestation of prosodic convergence between pairs of Italian speakers involved in a non-competitive game. Results show evidence of prosodic convergence and/or divergence between partners, where prosodic parameters and coordination strategies involved can vary across dialogue pairs. Also, degree of asymmetry in prosodic convergerce appears to be related to speaker empathy.
Keywords: Prosodic convergence, game dialogues, Italian, Big Five Questionnaire

Introduction
Conversational partners have been observed to adapt each other's speech over the course of the interaction. This phenomenon, variously termed as convergence, entrainment, alignment, accomodation, or adaptation, is considered crucial for mutual understanding and successful communication, and influenced by linguistic, social, and cultural factors (e.g. Giles et al. 1973). Research has been devoted to measuring prosodic parameters involved in speech adaptation in a number of languages (e.g. Levitan et al. 2015) not including Italian. This paper offers a preliminary investigation on prosodic convergence in Italian, and explores the influence of speakers' personality traits on the convergence process.

Method
Corpus
Our corpus consists of five dialogues where pairs of players are involved in the old Chinese Tangram Game, as developed within the PAGE project. Participants in a game round were given Tangram figures according to their role: the Director, who received a set of four Tangram figures, one of which marked by an arrow; the Matcher, who was given one of the figures belonging to the Director's set. Players could not see each other's figures, and goal of the game in each round was to establish whether the figure given to the Matcher was the same as the one marked by the arrow in the Director's set, or not. The game session consisted of 22 rounds, with an average duration of 30 min.

Speakers were selected according to a number of parameters which could influence adaptation, namely age, gender, and familiarity. They were all young adult females (aged 21-24), and MA student classmates. After the recording sessions, participants were administered the Big Five Questionnaire (BFQ-2, Caprara et al. 2007), a protocol used in psychology for assessing individual "Big Five Personality Factors" (Energy, Friendliness, Conscientiousness, Emotional Stability, Openness) along with their subdimensions.

Annotations and prosodic measurements

All dialogues were orthographically transcribed, and speech signal was manually annotated (by using Praat, Boersma 2001) along the following tiers: 1) Game rounds; 2) Inter-Pausal Units (speech bounded by silence longer than 100msec); 3) words; 4) syllables. The following prosodic parameters were automatically calculated via Praat scripts: pitch range (F0max-F0min, Hz), pitch level (F0 median, Hz), loudness (intensity, dB) and articulation rate (# syll/sec).

Measuring convergence

Given the explorative nature of this study, we started focussing on global aspects of speech coordination, i.e. those referring to similarity process undergoing at the level of the whole dialogue. We basically follow the approach proposed by Eldlund et al. (2009) in defining similarity as underlined by a) convergence, the process by which conversational partners' speech features become more similar over time until they converge; b) synchrony, when speakers' speech happen to have similar patterns over time. Due to space limitations, in this paper only results on convergence are presented. We looked for evidence of convergence by identifying cases in which speakers mean values were more similar to each other later in the dialogue. Accordingly, we splitted each game session into two halves: a window consisting of rounds 1-11 vs another window including rounds 12-22. Within each of the two windows, we compared mean values of speaker1 vs speaker2. Mean values found as significantly different in the first half but not in the second half were considered as evidence for convergence. Note that convergence can be realised on the opposite direction as a complementary manifestation of adaptation, i.e. divergence (Healy et al. 2014). Consequently, in our hypothesis mean values found as not significantly different in the first half but significantly different in the second half of the session were considered as evidence of divergence. All other cases were not taken as evidence for convergence or divergence.

Results

Prosodic convergence/divergence

Table 1 shows results of speaker1-speaker2 mean values comparison for each prosodic parameter, in the first vs. second halves of each game session. We found statistical evidence of convergence and/or divergence in four out of five dialogues: speakers in dialogue PZ become more similar in their voice loudness in the second part of the dialogue (convergence), whereas speakers in dialogue RC show complementary convergence by significantly diverging in their articulation rate in the second half of the session. In dialogues DS and CD we found both types of manifestation of overall coordination: participants in dialogue DS converge in their articulation rate, and diverge in the loudness of their voices, whereas speakers in dialogue CD converge in pitch range and diverge in pitch level. Speakers converging by some speech features yet diverging by some others in the same interaction has been reported (e.g. Bilous & Krauss, 1988, Eldlund et al. 2009).

Table 1. Comparison of speaker1 vs. speaker2 mean values in the first vs. second halves of dialogue (two-tailed t-test, t values only when significant: *=p<.05, **=p<.01, ***=p<.001). Light gray shaded boxes indicate convergence; dark gray shaded ones indicate divergence.

Dialogue	**Convergence/Divergence** speaker1-speaker2 mean values comparison, 1 half vs. 2 half of dialogue							
	Artic. rate		Pitch range		Pitch level		Loudness	
	1st half	2nd half	1st half	2nd half	1st half	2nd half	1st half	2nd half
CD	n.s.	n.s	2.18*	n.s.	n.s.	4.18***	2.29*	2.58*
DS	3.21**	n.s.	2.14*	2.16*	n.s.	n.s.	n.s.	2.16*
PP	n.s.	n.s.	n.s.	n.s.	-8.27***	-4.94***	4.66***	7.10***
PZ	n.s.	n.s.	n.s.	n.s.	-10.46***	-6.71***	-3.52**	n.s.
RC	n.s.	-2.69*	n.s.	n.s.	n.s.	n.s.	4.88***	4.89***

Prosodic convergence and speaker empathy

We explored the possible influence of speaker empathy (a subdimension of Friendliness) on the convergence process. We determined the degree of asymmetry in converging/diverging of each speaker in a pair by measuring mean values differences between the first and the second halves of the dialogue (only for the prosodic parameters involved). In Table 2, speakers whose absolute values were either greater (implying that they "converged more"), or smaller (implying that they "diverged less") with respect to their conversational partners, are the ones who consistently exhibit the higher scores for empathy in the pair (at least 10 T scores, 1 s.d.).

Table 2. Mean values differences (2^{nd}–1^{st} halves of dialogue) for each speaker in dialogues where convergence and/or divergence were observed, along with individual T scores for "Empathy" as assessed by the BFQ-2.

Dialogue	Speaker	Convergence 2^{nd}-1^{st} halves (mean values)	Divergence 2^{nd}-1^{st} halves (mean values)	Empathy (BFQ-2 T scores)
CD	sp1	10.12	9.50	58
	sp2	18.31	-2.50	70
DS	sp1	0.01	-0.90	56
	sp2	0.46	0.49	65
PZ	sp1	-0.04	-	59
	sp2	-2.53	-	76
RC	sp1	-	-0.43	61
	sp2	-	0.03	72

Discussion and conclusions

Results of this explorative study indicate that, at the whole dialogue level, Italian speakers tend to adapt their speech through a variable number of prosodic parameters, and by using different coordination strategies. These results are in line with those reported on languages investigated so far. In our data, we also observed that degree of asymmetry in convergerce/divergence appears to be related to speaker empathy. Though very preliminary, such observations are encouraging for future directions of our research.

References

Bilous F.,R., Krauss, R.M. 1998. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. Language and Communication 8, 183-194.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. Glot International 5(9/10), 131-151.

Caprara, G.V., Barbaranelli, C., Borgognani, L., Vecchione, M. 2007. Big Five Questionnaire-2, Giunti: Firenze.

Eldlund J., Heldner M., Hirschberg J. 2009. Pause and gap length in face-to-face interaction. In: Proceedings of Interspeech 2009, 2779-2782, Brighton, UK.

Giles, H., Taylor, D.M., Bourhis R.Y. 1973. Towards a theory of interpersonal accomodation through speech. Language in Society 2, 177-192.

Healey P., Purver M., and Howes C. 2014. Divergence in dialogue. *PloS one* 9(6) e98598, 1-6.

Levitan, R., Benus, S., Gravano A., Hirschberg J. 2015. Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. In Proceedings of SIGDial 2015, 325-334, Prague, Czech Republic.

PAGE (Prosodic And Gestural Entrainment in conversational interaction in diverse languages) project: http://page.home.amu.edu.pl/