Measuring prosodic entrainment in Italian collaborative game-based dialogues

M. Savino[1] , L. Lapertosa[1], A. Caffò[1], M. Refice[2]

[1]*Dept. of Education, Psychology, Communication, University of Bari, Italy*

[2]*Dept. of Electrical and Information Eng., Polytechnical Univ. of Bari, Italy*

SPECOM 2016 – Budapest, 23-27 August

Background

- Speakers tend to sound more similar over the course of interaction → convergence, adaptation, alignment, entrainment, coordination
  - Crucial for mutual understanding & successful communication, influenced by many factors (linguistic, social, interpersonal, cultural,…)
- Modelling speech adaptation also crucial for improving naturalness in voice-based human-machine interaction

Background & Aim

- Prosodic entrainment (prosodic-acoustic param) ➔ Studies on a number of languages (varieties of English, Swedish, German, Japanese, Spanish, Chinese, Slovak …) but NOT Italian
- Aim of this study ➔ preliminary contribution in filling this gap
- Explorative investigation on prosodic adaptation between Italian conversational partners

Corpus - Interaction paradigm

- Pairs of players involved in a collaborative game ➔ adaptation of Tangram Game (from PAGE project)

Director

Matcher

- Each game dialogue = 22 Tangram sets = 22 “Rounds”
- Players alternate role D/M in every Round
- Average duration of game sessions= 30 min
- With/without eye contact

Corpus - Speakers

- Speakers selected according to gender, age, familiarity
  - All parameters which could influence entrainment
- Twelve participants (six pairs)
- All females, aged 21-24, undergraduate classmates
- Also, all speakers coming from the same geolinguistic area (Bari)

Speech signal annotations

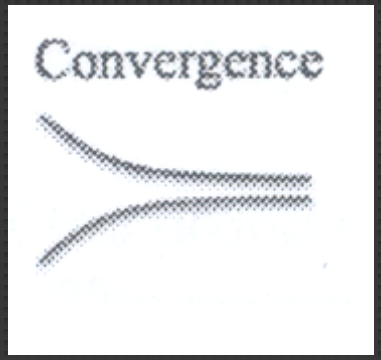
- Tangram Game Rounds
- InterPausal Units (silence > 100msec)
- Words
- Syllables

Speech signal all manually annotated (Praat)

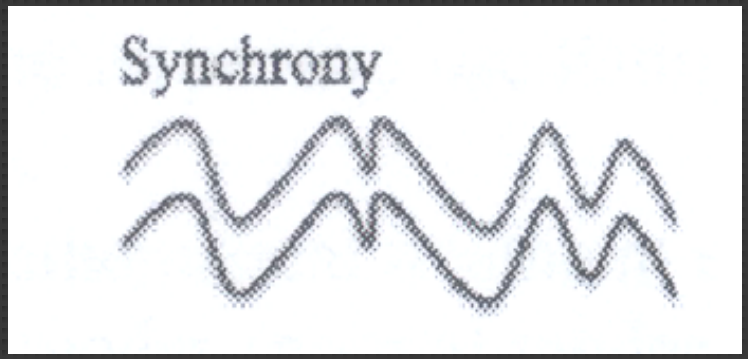
Prosodic measurements

- Fo range (Fomax-Fomin)
- Fo level (Fo median)
- Intensity
- Articulation rate (#syll/sec)
  - Automatically extracted (Praat scripts)
- In this study, measurements only on eye-contact condition data

Similarity processes (at dialogue level)



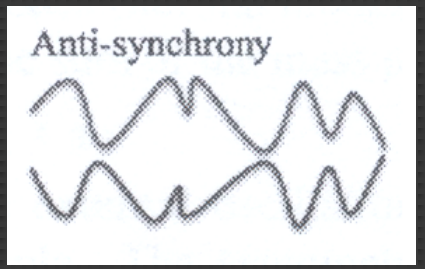
Speakers' speech features become more similar until they converge



Speakers' speech features show similar patterns

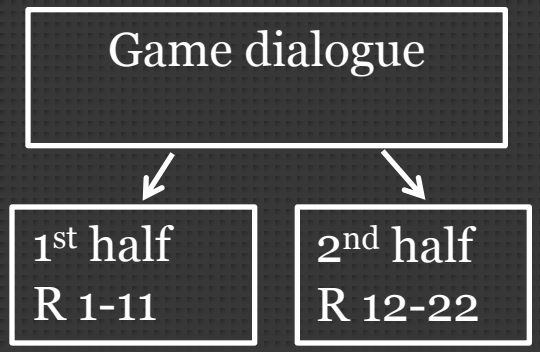
(Edlund et al. 2009, De Looze & Rauzy 2011)

- Not necessarily co-occurring
- Complementary manifestations also possible:



Similarity measurements

1)



Comparison (t-test) speaker1 vs speaker2 mean values:

- Different 1st half – not different 2nd half
  ➔ Convergence
- Not different 1st half – different 2nd half
  ➔ Divergence

2) Pearson's correlation speaker1-speaker2 mean values (Round) over the whole dialogue:

- Positive correlation ➔ Synchrony
- Negative correlation ➔ Anti-Synchrony

Results – Convergence & Synchrony

Convergence / Divergence

dialogue	Artic. rate		F0 range		F0 level		Intensity	
	1st half	2nd half	1st half	2nd half	1st half	2nd half	1st half	2nd half
BV	-3.73**	-3.97***	-2.33*	-2.34*	-6.42***	-9.35***	6.63***	8.75***
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***

Synchrony / Anti-Synchrony

dial	Artic. rate	F0 range	F0 level	Intensity
BV	.053	-.346	.048	.219
CD	.034	.185	-.120	-.295
DS	.523***	.191	-.381*	-.071
PP	-.097	-.217	.452**	.425**
PZ	.465**	-.204	.177	-.053
RC	-.098	-.078	.401*	.047

Conclusions

- Italian conversational partners show to adapt their speech through a variable number of prosodic parameters
- Overall speech coordination strategies (convergence, synchrony) can vary across speaker pairs
- Results compatible with those reported for other languages → common basis for modelling prosodic entrainment in multilingual spoken dialogue systems

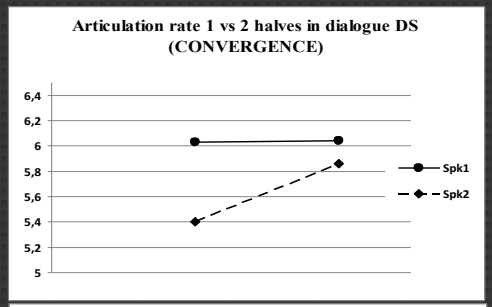
Thank you for your attention!

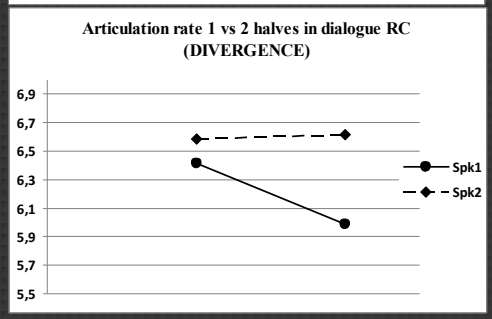
Entrainment & Personality factors

- After game sessions, participants were administered the Big Five Questionnaire (BFQ-2) ➔ assessing "Big Five" Personality Factors:
  - Energy, Friendliness, Conscientiousness, Emotional Stability, Openness (+ subdimensions)

Results - Convergence & Spkr Empathy

		Convergence 2nd-1st hal.	Divergence 2nd-1st hal.	Empathy (BFQ-2)
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





Partners who «converge more» / «diverge less» are the more empathic in the pair ➔ at least 1 sdev difference in BFQ-2 T scores for Empathy (subdimension of Friendliness)