

Amedeo De Dominicis

On conversation

(doi: 10.1418/87004)

Lingue e linguaggio (ISSN 1720-9331)

Fascicolo 1, gennaio-giugno 2017

Ente di afferenza:

()

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.
Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

L'articolo è messo a disposizione dell'utente in licenza per uso esclusivamente privato e personale, senza scopo di lucro e senza fini direttamente o indirettamente commerciali. Salvo quanto espressamente previsto dalla licenza d'uso Rivisteweb, è fatto divieto di riprodurre, trasmettere, distribuire o altrimenti utilizzare l'articolo, per qualsiasi scopo o fine. Tutti i diritti sono riservati.

ON CONVERSATION

AMEDEO DE DOMINICIS

ABSTRACT: What is the function of a dialog? Traditionally its function is to reach an agreement on a range of knowledge to be conveyed, in order to make this shared. The language, that is a grammar shared among the interlocutors, is just the main means by which this aim is pursued.

The paper investigates to what extent the metrics of speech can be induced by the pragmatic conditions of communication. The qualitative and statistic analysis demonstrates that rhythmical patterns vary according to the conversational goals of the speakers.

Results are compatible with a different answer to the question with which we began. The dialog function seems mainly to realize an activity together, and such final objectives are achieved through the instrument of a mutual understanding: in other words, the dialog seems above all a cooperation, which serves the language, but which does not aim for the language.

KEYWORDS: rhythm, conversation, synergy, pragmatics, Italian.

1. INTRODUCTION

According to some recent proposals (Fusaroli et al. 2012; Fusaroli, Raczaszek-Leonardi & Tylén 2014; Fusaroli & Tylén 2015), we make a specific distinction between classical and interpersonal approaches to the dialog. The classical approaches to dialog are largely based on models of individual linguistic processing. According to these approaches, each interlocutor enters the conversation as an independent entity, and brings his/her (linguistic and encyclopedic) competence in the dialog. In short, the conversation exists *after* the interlocutors, and it mirrors their competences. Thus, the interlocutors preliminarily have to share a common grammar, and that is the main reason why the dialog has been considered as a means for eliciting the grammar of a language through the so-called *map-task* method (cf. Brown et al. 1983; Anderson et al. 1991; Carletta et al. 1995; Carletta et al. 1997). According to these approaches, the function of the dialog is to make interlocutors reach a mutual understanding by means of a common grammar. The assumption is that the dialog is an interaction of independent and *internally* controlled individuals making the alignment of initially diverging situation models the key step for the success of their interaction. In addition, a second assumption is that the interaction is

even more *happy* as it brings to *align* the initial linguistic and cognitive differences of the interlocutors.

The main problem with the expectations of the classical approach is their incompatibility with two kinds of data: the stability or complementary nature of the dialog, and the synergic nature of the linguistic interaction between the interlocutors. As for the stability, many studies have revealed that the dialog is an instance of an orderly behavior among the parts of an assembly, in which the interlocutors realize their lexical, prosodic, and rhythmic (i.e. the distribution of their pauses) choices in a complementary way. This stability of the dialog clashes with the expectation of a dynamics bringing the interlocutors from an initial misalignment to a final alignment by means of the dialog.

Likewise, for the synergy between the interlocutors, here we will show that even in the case of a conflictual conversation, the interlocutors reciprocally align their linguistic behavior in terms of syllable and metrical feet durations (i.e. facts that are likely to be below their threshold of awareness), and they do that from the beginning of their interaction, up to contradicting the expected dynamic nature of the conversation.

According to the interpersonal approaches, language is a pre-eminent case of joint action (Clark 1996), and the dialog is a functional synergy, that pushes the interlocutors to reciprocally align their linguistic behaviors and representations on multiple levels. As for the function of such synergy, it responds to the environmental demands imposed on a dyad, to be in a joint task, to establish or maintain the social relations inside the group. It is not easy to say for what purpose the individuals in a group need to cooperate, probably, this is because it is necessary to express their membership in a group and then establish and maintain cohesive social relations. In any case, the dialog presents the same characteristics and functions of any other human action intended for a co-operative end; the interlocutors operate as the coordinated movements of the arm muscles that strike a hammer to drive a nail into the wall. The dialog seems to be predominantly a form of action, a special action because it is co-operative, but not different from those actions with a motor characteristic or with a social nature that are imposed on individuals as members of a group, and therefore because they need to express their membership using shared forms of acting. In short, according to this latter approach, the main reason why we engage a conversation is to realize an activity, and particularly an activity together, by means of a mutual understanding. The mutual understanding is no more the main goal of the conversation, but it is a simple means through which the dyad reaches the actual goal of a conversation, and realizes a joint activity.

2. ORIGINS AND CHARACTERISTICS OF THE INTERPERSONAL SYNERGETIC APPROACH

The notion of synergy originates from the study of movement, as a way of describing the functional coordination of multi-element systems (Bernstein 1967; Turvey 1977). The model originates from the motor control theory by Bernstein. He found it unlikely that the central nervous system could be able to finely control all the possible movements (the degrees of freedom) of each single muscle individually to create coherently directed movements. Rather, he suggested that muscles form flexible function-specific self-organizing assemblies, i.e. synergies, by locally coupling and constraining each other's degrees of freedom, greatly reducing the amount of control needed. For instance, if we want to strike a chisel with a hammer, the variability of the trajectory of the tip of the hammer across a series of strikes is smaller than the variability of the trajectories of the individual joints on the hammering arm (Bernstein 1967).

An example of such a synergy can be the motor coupling of two agents cooperating in order to achieve a common goal: if two agents are carrying a big piece of furniture and one of them raises one side too high up, the other, to keep the piece balanced, can compensate by raising the opposite side in proportional ways. In this sense, their actions come to complement each other, for example, adjusting in opposite directions, maintaining the furniture balanced. Similarly, interlocutors become interdependent in their linguistic behavior.

A recent analysis of the trajectories of flocks of starlings may be another good example of a synergetic behavior driven by the positional adjacency: the birds located at the border of the flock trigger and lead the fluctuation of the formation. The analyses by Giorgio Parisi and his team (Attanasi et al. 2014a, 2014b) reveal that in flocks of starlings the velocity fluctuations of two distant birds mutually influence each other, and that the seemingly disorganized swarms have some sophisticated features: i.e. they are a self-organized (synergetic) whole, and coordinate their behavior so as to optimize their ability to react collectively (e.g. to avoid predators or to attract sexual partners).

3. PREVIOUS WORKS

The interpersonal approach has been developed in recent works (Konvalinka et al. 2011; Fusaroli et al. 2012; Fusaroli, Raczaszek-Leonardi & Tylén 2014; Fusaroli & Tylén 2015).

Konvalinka et al. (2011) analyze a Spanish fire-walking ritual (in San Pedro Manrique), and give an account for the so-called collective arousal effect. They measure heart rates (from active participants and spectators) during the course of the fire-walking ritual. Thus they investigate the synchronized arousal instead of the synchronized body movements. They also account separately for the measures referring to spectators that are related to the fire-walkers, and spectators that are not (observers). They find that the same heart pace pattern occurs for both the fire-walkers and their related spectators, whose heart rates peaked for the walk of their relatives and friends. But this pattern is not shared with the unrelated observers. Thus what these measures show is not just that the heart rates go up at the time of the fire-walk, but that there are shared patterns of arousal that span the fire-walking ritual across all of the individual fire-walks, between performers and spectators that are related or tangentially related. Moreover, they show that there is a social catalyzer that promotes a shared pattern of arousal inside the borders of a given social group, whereas it inhibits the same sharing across a social border. A dialogue is a special kind of social catalyzer; by means of the dialogue people share and coordinate their behavior, in order to carry out a synergy.

Fusaroli et al. (2012) conduct an experiment concerning the copy effects between the interlocutors' speech turns in a conversation and the benefits it brings to solving a perceptual problem. The experimental expectation is that the occurrence of the copy effects (at lexical, prosodic, and rhythmic levels) is compatible only with the model of the dialog as an interpersonal synergy. Previous researches had shown that interlocutors tend to adapt to each other's lexical choices (Fusaroli et al. 2012; Garrod & Anderson 1987; Krauss & Weinheimer 1966), prosodic patterns (Kousidis & Dorrán 2009; Levitan & Hirschberg 2011; Pardo 2006; Pardo et al. 2012), and rhythms of speech and pauses (McFarland 2001; Wilson & Wilson 2005). According to these findings, the conversation is a coherent, co-constructed text to assess structures of recurrence of prosody, speech/pause, and lexicon. In their experiment dyads of participants cooperate linguistically to solve a perceptual task (the participants have to individually indicate which of the 2 displays contains a contrast oddball) and then they have to discuss about it. The experimental results confirm that the degree of global linguistic convergence of confidence expressions (such as 'to be sure', 'to see', 'to think', 'to know', 'to guess', 'to believe', 'to shoot') significantly correlates with collective benefit.

4. RHYTHM MEASURES

According to the interpersonal and synergetic approach to the conversation, we present some experimental verifications which show how and to what extent the speech rhythm can be induced and shared by pragmatic conditions, and by the interpersonal synergy in a conversation.

Two main approaches to linguistic rhythm exist in the literature: the hypothesis of rhythmic discrete types and the assumption of rhythm as a variable property that does not belong to the linguistic system, but to conversational interaction. In this latter approach the rhythm function is to handle cooperation and conflict among the speakers. Therefore it is not stable, but varies according to its conversational functions.

4.1 Rhythm-property of the system

The hypothesis of rhythmic types goes back to the forties (Lloyd James 1940; Pike 1945; Abercrombie 1967; Faure, Hirst & Chafcouloff 1980; Dauer 1983). It mainly consists of a binary classification (syllable-timed/stress-timed languages¹). But it has not yet been clearly experimentally validated (e.g. Shen & Peterson 1962; Bolinger 1965; O'Connor 1965; Uldall 1971; Lea 1974; Lehiste 1977; Donovan & Darwin 1979; Roach 1982; Wenk & Wiolland 1982; Borzone de Manrique & Signorini 1983; Dauer 1983; Drake & Palmer 1993). According to a weaker hypothesis, rhythm is a perceptual impression arising from the convergence of some clusters of phonological properties typical of a given language (e.g. Dasher & Bolinger 1982; Nespor & Vogel 1986; Dauer 1987; Bertinetto 1981, 1989; Nespor 1990; Ramus, Nespor & Mehler 1999). The linguistic typology (syllable/stress-timed) is not discrete and different systems are spread out over a *continuum*.

Also according to the PVI hypothesis (Low & Grabe 1995; Low, Grabe & Nolan 2000; Grabe & Low 2002; Patel & Daniele 2003), rhythm is an intrinsic property of the system.

4.2 Rhythm-variable property of conversation

This hypothesis derives from conversational analysis studies, and represents rhythmic features in Gestalt terms. Recently, a new impetus has been given by the so-called phonetic-details studies (cf. Sacks, Schegloff & Jefferson 1974; Erickson 1982; Erickson & Shultz 1982; Cutler 1991; Couper-Kulhen 1989, 1990, 1993, 2001; Buder 1986, 1991, 1996; Auer et al. 1999; Buder &

¹ In syllable-timed languages the duration of every syllable is equal. In stress-timed languages the interval between two stressed syllables is equal.

Eriksson 1997, 1999; Local 2003; Fon 2006; House 2007; Russo & Barry 2008; Arvaniti 2009; Reed 2010). In this paradigm, during interaction, rhythm may vary due to the conversational tasks, it is not a property of the system, but a tactical resource of the speaker.

To this paradigm three more branches belong: the studies on the metrical feet variability (*heterometry*) (Brown & Weishaar 2010); the studies on rhythm as an entrainment phenomenon (Cummins & Port 1998; Port 2003; Cummins 2009); the studies of rhythm as an Adaptive Oscillator (Port, Cummins & Gasser 1996).

5. EXPERIMENTAL ANALYSIS

Before starting the main analysis of the polemical corpus, we ran a preliminary experimental test in order to verify to what extent the speech metrics can be induced by some pragmatic conditions. Two corpora have been elicited. The first corpus was obtained by an experimental collaborative task in which the subjects were asked to synchronize their speech with a recorded one. The second is a natural corpus in which two speakers are engaged in a polemical interaction (the so-called quarrel between Vittorio Sgarbi and Mike Bongiorno during the Italian TV show *Telemike* in 1991). In each of the two corpora we took two measurements: the interstress intervals (henceforth ‘Acc’ = the temporal distance between the stressed syllables); the syllabic intervals (henceforth ‘Syl’ = the duration of stressed and unstressed syllables). These measurements were used to check the metrical typology (stress/syllable-timing) and its variation along the corpus.

5.1 *Experiment on the collaborative corpus*

We recorded the sentence *Il capostazione ha spento la luce* (‘The station master has turned the lights out’). The rhythm of the original signal (A) is not syllable-timed, nor stress-timed (the duration of Acc in the original signal A is not constant, and the same goes for the duration of Syl). Then we manipulated the signal A in order to build new ones with constant Syl or Acc (signals D and E). On these signals we built a *Listen & Repeat* test. The working hypothesis was that listening to these signals (B, C, D, E) will induce the listener to a syllable-timing or a stress-timing rhythm, according to the manipulated signals. To the same purpose, before the original signal A, we inserted three beeps,² 150 ms apart (equal to the mean Syl in the original signal) and 496 ms

² At least three evenly spaced beats are required in order to establish an isochronous chain (Couper-Kuhlen 1990: 16).

apart (equal to the mean Acc in the original signal). Thus we produced the signals B and C. We also manipulated the duration of Syl and Acc in order to equalize their duration respectively to 150 ms and to 496 ms. Thus we produced the signals E and D. Both mean values (150 ms and 497 ms) are only artificial targets. They do not represent an actual target rhythm, or a natural voice to be imitated: they are just an experimental tool to “induce” a rhythm by a subject. The signals for this listening test (PASSIVE CORPUS) are listed in Table 1.

SIGNAL	
PASSIVE CORPUS	
A	the natural original signal (writer's voice)
B	three beeps (150 ms apart) + signal A
C	three beeps (496 ms apart) + signal A
D	signal A: equalized Acc durations: 496 ms
E	signal A: equalized Syl durations: 150 ms
ACTIVE CORPUS	
1	the natural original signal (subject's voice)
2	signal recorded after listening to the signal B
3	signal recorded after listening to the signal C
4	signal recorded after listening to the signal D
5	signal recorded after listening to the signal E

TABLE 1. PASSIVE AND ACTIVE CORPORA.

The ACTIVE CORPUS is composed of the sentences that the subjects recorded after listening to the passive corpus.

Five university students took part in the experiment: S1 (male, age 52, born in Rome where he lives); S2 (female, age 24, born in Terni where she lives); S3 (female, age 51, born in Orune-Nu, but living in Tuscania-Vt); S4 (female, age 19, born in Alatri-Fr where she lives), S5 (female, age 19, born in Civitavecchia where she lives). They listen to signal A, and are asked to utter the same sentence into the microphone (signal 1 of the active corpus). Then they listen to the further 4 signals and repeat them into the microphone, trying to imitate them and keep as close to the same timing of the signal they listened to with headphones. Thus, the ACTIVE CORPUS contains 25 signals (5 per subject), as shown in Table 1.

5.2 Experimental expectations and results

Compared to signal 1 (recorded at the beginning of the session), signals 4 and 3 should show equalized Acc durations and close to 496 ms (but the absolute value depends on the speaking rate). Signals 5 and 2, should show equalized Syl durations and close to 150 ms (but the absolute value depends on the

speaking rate). If these expectations are confirmed, then the syllable-timed or stress-timed rhythm is an effect of the communicative interaction. As we are able to induce, it is not a property of the linguistic system. The results validate these expectations. Table 2 shows the Syl and Acc durations in the active corpus, and their standard deviation (σ). Signals 2 and 5 systematically approach the reference value (150 ms) as compared to signal 1; likewise, 3 and 4 systematically approach the reference value (496 ms). Therefore the σ decreases.

SYLLABLES		SIGNAL 1		SIGNAL 2		SIGNAL 5	
		ms	σ	ms	σ	ms	σ
	S1	141.2	43.78	159.8	47.44	149.0	36.38
	S2	201.7	82.06	197.5	47.07	190.3	46.42
	S3	176.3	74.82	156.4	57.47	156.5	50.17
	S4	177.3	55.12	163.9	35.78	152.1	35.70
	S5	144.8	44.82	152.7	36.56	149.2	23.50
INTERSTRESS INTERVALS		SIGNAL 1		SIGNAL 3		SIGNAL 4	
		ms	σ	ms	σ	ms	σ
	S1	402.3	78.61	476.0	76.21	474.0	53.69
	S2	620.0	253.74	585.3	110.88	550.3	152.16
	S3	560.6	232.02	493.0	124.38	500.3	87.06
	S4	510.6	115.86	444.0	46.16	477.0	18.68
	S5	418.3	65.54	477.3	29.29	479.6	23.09

TABLE 2. MEAN SYL/ACC DURATIONS (MS) AND THEIR STANDARD DEVIATION (σ).

5.3 Experiment on the polemical corpus

As for the polemical corpus it is the quarrel between two Italian TV showmen: Mike Bongiorno and Vittorio Sgarbi (*Telemike* in 1991). It was downloaded from *YouTube*. Its low audio quality creates no problem with the duration measurements. However, the low quality of that material, the occurrence of several overlappings, and the compressed format of *YouTube* audio (mp3) impede any spectral analysis or any extraction of F0. The polemical corpus is a communicative situation where the speakers do not collaborate, but manage to hinder and sabotage each other. A typical expression of this hostile intent is represented by repetitions, interruptions, and overlappings. The audio recording has been transcribed (both in IPA, and in orthographic layout); then we extracted the sequences of speech turns containing overlappings, repetitions, interruptions, and among them we selected those which could be analyzed (the choice was made taking into consideration the sequences without environmental noises, clappings, yells, etc.); these sequences of speech turns are called *chains*. There are 23 chains (1-23) in the quarrel. Each chain repre-

- ³ Overlapping turn-taking is represented by columned text.

(‘this is the idea’)

- (16) Mike: *dici tu, va bene?* *io non dic nessuna cazzata*
 (‘you say, ok?’) (‘I’m not blundering’)
 Sgarbi: *sì, d’acco le diciamo insieme siam* *insieme a dirlo io*
 (‘yes ok, we say them together, we both’) (‘say it, I’)
- (17) Mike: *adesso parlo io sì adesso mettila giù*
 (‘now it’s up to me to speak, yes now put it down’)
 Sgarbi: *ora litiga vuoi far a pugni con me no puoi parlare no*
 (‘now fight! do you want’) (‘to come to blows with me? No you can’t say anything’)
- (18) Mike: *no perché tu devi smettere*
 (‘no, because you must stop’)
 Sgarbi: *voglio vuoi fare una rissa no*
 (‘I want do you want to fight with me? Don’t you?’)
- (19) Mike: *smetti*
 (‘stop’)
 Sgarbi: *tu stai dicendo delle cose che non sai*
 (‘you are talking...’) (‘about what you know nothing of’)
- (20) Mike: *io non ho ancora io non ho*
 (‘I have not yet, I have not’)
 Sgarbi: *quelli che parlano a vanvera è andato*
 (‘those who continuo talking nonsense, it is gone’)
- (21) Mike: *n’ [non] ho ancora finito la mia frase ...vuoi tacere?*
 (‘I have not finished my sentence yet ... do you want stop talking?’)
- (22) Mike: *parli sempre taci qualche volta*
 (‘you never stop talking’) (‘stop talking sometimes’)
 Sgarbi: *mi fai parlare mi fai parlare*
 (‘you are talking...’) (‘let me talk’) (‘let me talk’)
- (23) Mike: *taci qualche volta*
 (‘stop talking sometimes’)

The signals are annotated by means of 6 *Praat Tiers* (see Figures 1-4 and 6), as follows: (1) orthographic transcription: Mike, (2) orthographic transcription: Sgarbi, (3) Mike’s Syl: IPA and boundaries, (4) Sgarbi’s Syl: IPA and boundaries, (5) Mike’s Acc: IPA and boundaries, (6) Sgarbi’s Acc: IPA and boundaries.

5.4 Experimental expectations and results

In the polemical *corpus* we expect a minimal degree of rhythmic integration: i.e. anisochrony. The results confirm these expectations. Indeed, no stable

rhythmic pattern exists. Furthermore, the metric of each turn changes according to the conversational purposes; in particular, the speaker may borrow his interlocutor's rhythm – or implement the opposite (for instance a syllable-timed vs. a stress-timed rhythm) – in order to dominate by cutting or easing the antagonist's rhythm. The chaining and the syntagmatic succession of metrical types in each speech turn is a function of the conversational strategy of the speaker to create dominance.

STRAT- EGY	CHAIN	MIKE	SGARBI	SYLLD	σ (SYLLD)	ACCD	σ (ACCD)	T	ANO
SKIR- MISH	1	A		151.5	52.15	266.1	33.20		
	2		S	110.6	22.03	272	107.25		
	3		S	121.5	17.67	440.5	228.39		
	4		A	100.8	44.25	254.8	20.53		
	5		A	108.9	52.79	225.4	15.80		
	6		S	164	4.24	244	74.95		
	7		S	319	2.82	478.5	144.95		
	8		S	392	11.31	467	224.86		
	9		A	150.1	85.41	538.5	10.60		
	10	S		117.3	6.02	1 stress	--		
	11		A	187.8	106.39	540	14.14		
	12		A	77	29.17	172.2	16.91		
	13	S		120.6	8.96	297.3	35.83		
TRUCE	14		*	114.6	53.01	190.6	94.11		
	15		*	156.2	47.78	249.5	149.19		
MI- METIC	16	Ac	Ac	Mike: 180.4 Sgarbi: 144.3	Mike: 43.8 Sgarbi: 45.7	Mike: 376 Sgarbi: 345.4	Mike: 14.4 Sgarbi: 27.2	0.07	0.08
ROLL- ING	17	*	*	Mike: 126.8 Sgarbi: 128.3	Mike: 56.5 Sgarbi: 80.7	Mike: 346.2 Sgarbi: 217.6	Mike: 47.7 Sgarbi: 102.5	0.01	0.02
INTER- VAL	18	A		168.2	63.38	348.2	11.67		
	19		S	144	13.1	302.7	166.44		
MI- METIC	20	Sc	Sc	Mike: 181.7 Sgarbi: 166.5	Mike: 22.09 Sgarbi: 36.33	Mike: 566 Sgarbi: 494.3	Mike: 90.50 Sgarbi: 217.19	0.50	0.50
END	21	A		168.1	73.82	401	59.99		
	22	A		195.4	62.51	359	41.76		
	23	S		263.6	27.5	362	76.36		

TABLE 3: SEQUENCE OF TURNS AND RHYTHM TREND. A = STRESS-TIMED; S = SYLLABLE-TIMED; * = ASYNCHRONOUS RHYTHM; C = SYNCHRONIZED DURATIONS; SYLLD = SYLLABLES MEAN DURATION; ACCD = INTERSTRESS INTERVALS MEAN DURATION; σ = STD DV; T = TWO SAMPLE STUDENT'S T-

TEST, ASSUMING AN UNEQUAL VARIANCE & A LEVEL OF CONFIDENCE $\alpha = 0.05$; ANO = ONE-WAY, OR SINGLE-FACTOR ANOVA ($\alpha = 0.05$). NON-SIGNIFICANT VALUES ARE ON A GREY BACKGROUND.

As is shown in Table 3, at the beginning both speakers alternate different metrics, in a sort of ‘skirmish’ (opposed rhythm speech turns). A ‘truce’ follows: Sgarbi tries to disrupt the metrical strategy of Mike, using an asynchronous rhythm. Then, both speakers resume their confrontation, but change their tactics: there is the first instance of speech turns overlapping (‘mimetic’) where their rhythm is common: stress-timed and synchronized. Next, is a second overlapping (‘rolling’) where, again, the metrics of the quarrellers are completely asynchronous, but dynamically tuned. Then, there are two ‘interval’ speech chains: a stress-timed trend chain by Mike and a following one by Sgarbi, showing an opposed syllable-timed trend. Then, the ‘mimetic’ tactic resumes, but the rhythmic features are reversed, as compared to the previous one: there is a third turns overlapping with a common syllable-timed and synchronized metric. In the ‘end’, three chains by Mike show an alternate rhythmical trend: stress- and syllable-timed. Four examples of these turns are given below.

1. TRUCE-RHYTHM (Table 3: chains 14-15; Figures 1-2). Sgarbi shows no rhythmic isochrony: an extreme case of polemical strategy (maybe in order to sabotage Mike’s rhythm). His rhythm is anisochronous: the σ values are very high both for the Acc and the Syl mean durations.

2. MIMETIC-RHYTHM (Table 3: chain 16; Figure 3). It is a first overlapping where both speakers tend to have a common stress-timed rhythm, and synchronized interstress boundaries: the difference between their mean Acc duration is not significantly different, as assessed by Student’s ($0.07 > 0.05$) and Anova tests ($0.08 > 0.05$).

3. ROLLING-RHYTHM (Table 3: chain 17; Figure 4), a second overlapping. The turns are anisochronous (the difference between their mean Acc duration is significantly different, as assessed by Student’s - $0.01 < 0.05$ - and Anova tests - $0.02 < 0.05$), but with a peculiarity: both speakers undertake a sort of rhythmic ‘rolling relay’, where the metrics of both quarrellers is *dynamically tuned*, i.e. each turn takes up the Acc durational trend towards the increasing or decreasing of the previous one, uttered by the interlocutor. As you see in Figure 5, Sgarbi produces a sequence of three increasing intervals (62-190-388 ms), followed by a reply by Mike with three equally increasing intervals (336-322-417 ms); then, Sgarbi reverses the trend, realizing a 170 ms interval and Mike pursues the decreasing trend with a 291 ms interval. Finally, Sgarbi reverses the trend again and produces a 236 ms interval, and Mike replies with the same increasing trend (365 ms).

4. MIMETIC-RHYTHM (Table 3: chain 20; Figure 6). In this third overlapping, both speakers tend to have a common syllable-timed rhythm and

they even tend to overlap, to make isotopic (synchronized) their syllabic boundaries: the difference between their mean Syl is not significantly different, as assessed by Student's ($0.50 > 0.05$) and Anova tests ($0.50 > 0.05$).

6. DISCUSSION

The rhythmic strategies are particularly interesting, because they can be argued to be presumably under the speaker's recognition threshold. The results of the preliminary experiment on the collaborative corpus simply show that the rhythmic patterns can be induced, and are not (or, at least, not entirely) imposed by the linguistic system. They rather pertain to the speakers' choices and performance. The results of the main experiment on the polemical natural corpus go further, and show that during a conversation the interlocutors' rhythm enters a sort of synergy. In this synergy the implementation of a given rhythm depends on the conversational strategy of each speaker, and on the demand to carry out a joint activity with his interlocutor. These findings are highly compatible with the model of conversation as an interpersonal synergy. And this synergy is also acting during a quarrel, that is in a conversation where the goal is not to reach a common understanding, not to share or transfer some information, but to overpower, to make the interlocutor unable to react.

Moreover, the data bring some consequences concerning the role of the conversation as a reliable means to elicit the grammar of a spoken language. With the expression "grammar of a spoken language" we are referring to the reconstruction of the linguistic system of an oral language. In fact, the conversation has been assumed to be representative of a (semi) spontaneous linguistic behavior, such as the linguistic interaction between two (or more) speakers engaged in a dialog. But, according to our findings, conversation is actually not a good witness of the grammar the speakers are using, or – at least – it does not make a very reliable witness. The reason is that the interlocutors' performance during a conversation – at least for their rhythmic choices – is not grammar-driven, but action-driven: they assume a rhythm or another according not to the language (and the grammar) they are speaking, but according to the interlocutor's rhythmic choices, to the demand to carry out a synergy. In short, conversation is a kind of joint action, locally motivated, more than an example of a particular realization of an abstract linguistic system, shared by both talkers.⁴

⁴ The grammar of a spoken language is not entirely the same thing as the rhythm of that language, but the rhythm is one of the main components of its prosody, and thus of its grammar. Prosody is a distinctive constituent of speech, and does not occur in written texts. Thus the rhythm and the prosody mainly shape the grammar of a spoken language.

One more consideration can be added. We argued before that a problem with the traditional approach to the dialog is that it predicts that the linguistic alignment between the interlocutors should be dynamic, that is it should not appear at the very beginning of a conversation. Whereas the interpersonal approach predicts that the alignment should happen all over the dialog. Our data seem not to fulfill the expectation of the interpersonal approach, because in the quarrel between Mike and Sgarbi (polemical corpus) the strong alignment happens in the middle of the conversation, whereas at the beginning and at the end each quarreler shows a personal and unshared rhythmic behavior: Mike starts with stress-timed rhythmic trend, and Sgarbi with a syllable-timed one; and Mike ends with the same stress-timing; in the middle of the quarrel-conversation various types of rhythmic alignments happen. But these apparent falsifications of the model can be interpreted as rhythmic events that serve as markers of the edges of a dialog. The reason why a dialog needs to be segregated from the outside, and delimited by an initial and a final border is that the dialog is a kind of ritual act. As every social routine, the dialog can be understood as a social plot in which agents (interlocutors) need to enter and leave. All stages of this plot comply with a *morphology* and syntagmatics, in the sense of Propp (1928). Thus the beginning and the end of the conversation, its edges, provide a structural path to and from the plot. In the beginning of the conversation-quarrel between Mike and Sgarbi each interlocutor keeps his rhythmic identity (Mike shows a stress-timed identity, and Sgarbi a syllable-timed isochrony), then both lose their identities and becomes interdependent in a new synergy, a coherent, co-constructed new entity: the conversation. At the end of the dialog, both interlocutors have to leave the dyad and come back to their previous identities: so Mike ends the conversation putting on the same stress-timed rhythm he was using before entering the dialog.

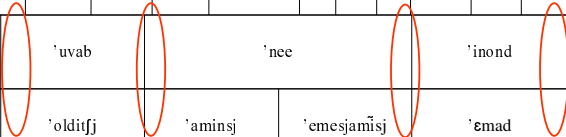
Mike: parole			•										
Sgarbi: no non puoi				dirle		perché		dici		delle		caz	
pa	'rɔ		le										
'nɔ	non		'pwɔj	'dir	le	per	'χe	'di	tʃi	'del	le	kaθ	
pa'rɔ													
'nɔ	non'pwɔ			j'di		rleper'χe			'di	tʃi'de		llekaθ	

FIGURE 1. SGARBI: *NO NON PUOI DIRLE PERCHÉ DICI DELLE CAZZ* ['NO, YOU CAN'T TATTLE THIS AS YOU TALK BULLSHIT'].

Sgarbi: questo	è il	concetto		
'kɛs	'tɛl	kon	'tʃɛt	to
k	'ɛst	'ɛlkontʃ		'ɛtto

FIGURE 2: SGARBI: *QUESTO È IL CONCETTO* ['THIS IS THE IDEA'].

Mike: dici			tu	va bene						io	non	dic	nessuna		caz								
Sgarbi: sì	d'acco			le diciamo			insieme		siam	insiem		a	dirlo			io							
'di		tʃi		'tu	va	'be	ne					'i	non	'di	kne	'su	na	kats					
	'sì	da	'ko	ldi	'tʃja	min	'sje	me	sja	mĩ	'sje	ma	'dir		lo				'i	o			
	'diʃfit				'uvab				'nee			'inond			'iknes					'unakats			
s		'idak			'olditʃj		'aminsj					'emesjamĩsj			'emad			'irlo					'io

FIGURE 3: MIKE: *DICI TU. VA BENE. IO NON DICO NESSUNA CAZ.* ['You say that. OK. I don't talk bullshit']; SGARBI: *SÌ D'ACCORDO, LE DICIAMO INSIEME, SIAMO INSIEME A DIRLO...IO* ['Yes, ok, we talk at once. We talk at once, ..I'].


					Mike: adesso				parlo				io		sì		adesso				mettila giù						
Sgarbi: ora		litiga			vuoi	far	a	pugni		con me						no				puoi		parlare		no			
					a	'des	so	'par		lo		'i	o	si	a	'dɛs	so	'mɛt		ti	la	'dʒu					
'o	ra	'li	ti	ga	'vwɔj	fa	rap	'puɲ	ɲi	kom	'mɛ					'no			'pwɔj		par	'la	re	'no			
							'essop				'arlo			'iosiad				'ɛssom		'ɛttɪladʒ					u		
'oral		'itigavw			'ɔjɸ		'arapp		'uɲikomm			'ɛ								'ɔpw		'ɔjparl			'aren		'o

FIGURE 4: MIKE: *ADESSO PARLO IO. SÌ ADESSO METTILA GIÙ* ['No it's my turn. Yes now put it down']; SGARBI: *ORA LITIGHIAMO. VUOI FARE A PUGNI CON ME? NO PUOI PARLARE. NO* ['Now we are going to have a quarrel. Do you want to box me? No you can talk. No'].
ME? NO PUOI PARLARE. NO ['Now we are going to have a quarrel. Do you want to box me? No you can talk. No'].

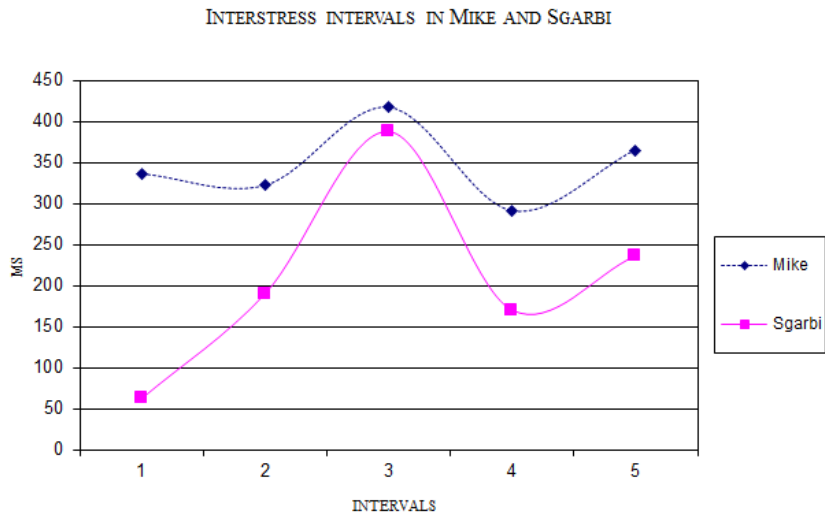


FIGURE 5: ROLLING-RHYTHM. VARIATION OF INTERSTRESS INTERVALS DURATION BY MIKE AND SGARBI, ARRANGED IN SEQUENCE.

Mike: io	non	ho	ancora				io			non		ho	
Sgarbi: quelli		che	parlano		a	vanvera				è	andato		
'i	o	non	ɔ	aŋ	'kɔ		ra		'i	o		nɔ	nɔ
'kwɛl	li	ke	'par	lno	a	'van	ve	ra		ɛ	an	'da	to
	'iononɔaŋk				'ɔra				'ionɔnɔ				
kw	'ellikɛp			'arlnoav								'ato	

FIGURE 6: MIKE: *IO NON HO ANCORA*., *IO NON HO* ['I HAVE NOT YET .. I HAVE NOT..']; SGARBI: *QUELLI CHE PARLANO A VANVERA. È ANDATO* ['THOSE WHO BLETHER. IT'S GONE']

REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Anderson, A., M. Bader, E. Gurman Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson & R. Weinert (1991). The HCRC map task corpus. *Language and Speech* 34(4). 351-366.
- Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica* 66(1-2). 46-63.
- Attanasi, A., A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri & M. Viale (2014a). Finite-Size Scaling as a Way to Probe Near-Criticality in Natural Swarms. *Physical Review Letters* 113. 238102.
- Attanasi, A., A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri & M. Viale (2014b). Collective Behaviour without Collective Order in Wild Swarms of Midges. *PLoS Computational Biology* 10(7).
- Auer, P., E. Couper-Kuhlen & F. Muller (1999). *Language in Time*. New York: Oxford University Press.
- Bernstein, N.A. (1967). *Coordination and regulation of movement*. New York: Pergamon Press.
- Bertinetto, P.M. (1981). *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Accademia della Crusca.
- Bertinetto, P.M. (1989). Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de Phonétique Appliquée* 91/93. 99-130.
- Bolinger, D. (1965). Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (eds.), *Forms of English: Accent, Morpheme, Order*, 139-180. Cambridge MA: Harvard University Press.
- Borzone de Manrique, A.M. & A. Signorini (1983). Segmental durations and the rhythm in Spanish. *Journal of Phonetics* 11. 117-128.
- Brown, G., A. Anderson, R. Shillcock, & G. Yule (1983). *Teaching talk: strategies for production and assessment*. Cambridge: Cambridge University Press.
- Brown, S. & K. Weishaar (2010). Speech is "heterometric": The changing rhythms of speech. In Hasegawa-Johnson, Mark et al. (eds.), *Proceedings of the Fifth International Conference on Speech Prosody*, paper 074. Chicago: ISCA.
- Buder, E.H. (1986). *Coherence of speech rhythms in conversations: Autocorrelation analysis of fundamental voice frequency*. Toronto: Toronto Semiotic Circle.
- Buder, E.H. (1991). *Vocal synchrony in conversations: spectral analysis of fundamental voice frequency*. PhD dissertation, University of Wisconsin, Madison, WI.
- Buder, E.H. (1996). Dynamics of speech processes in dyadic interaction. In J.H. Watt & C.A. VanLear (eds.), *Dynamic Patterns in Communication Processes*, 301-325. Thousand Oaks, CA: Sage.

- Buder, E.H. & A. Eriksson (1997). Prosodic cycles and interpersonal synchrony in American English and Swedish. In G. Kokkinakis, N. Fakotakis, E. Dermatas (eds.), *Eurospeech '97 Proceedings*, vol. 1, 235-238. Grenoble: European Speech Comm. Association.
- Buder, E.H. & A. Eriksson (1999). Time-series analysis of conversational prosody for the identification of rhythmic units. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, A. C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 2, 1071-1074. San Francisco: Univ. of California.
- Carletta J., A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, & A. Anderson (1995). The coding of dialogue structure in a corpus. In J. A. Andernach, Stan P. van de Burgt & Gerrit F. van der Hoeven (eds.), *Proceedings of the Ninth Twente Workshop on Language Technology: corpus-based approaches to dialogue modelling*, 25-34. Enschede: Universiteit Twente.
- Carletta, J., A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon & A. Anderson (1997). The reliability of a dialogue structure coding scheme. *Journal of Computational Linguistics* 23(1). 13-31.
- Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Couper-Kulhen, E. (1989). Speech rhythm at turn transitions: its functioning in everyday conversation. Part I, *KontRI*, Working Paper No. 5. Konstanz: Fachgruppe Sprachwissenschaft, University of Konstanz.
- Couper-Kulhen, E. (1990). Discovering rhythm in conversational English: Perceptual and acoustic approaches to the analysis of isochrony, *KontRI*, Working Paper No. 13. Konstanz: Fachgruppe Sprachwissenschaft, University of Konstanz.
- Couper-Kulhen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*. Amsterdam: Benjamins.
- Couper-Kulhen, E. (2001). Intonation and Discourse: Current Views from Within. In D. Schiffrin, D. Tannen, & H. Ehnberger Hamilton (eds.), *The Handbook of Discourse Analysis*, 13-34. Oxford: Blackwell.
- Cummins, F. (2009). Rhythm as Entrainment: The Case of Synchronous Speech. *Journal of Phonetics* 37(1). 16-28.
- Cummins, F. & R.F. Port (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26(2). 145-171.
- Cutler, A. (1991). Prosody in situations of communication: salience and segmentation. In *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 1, 264-270. Aix-en-Provence.
- Dasher, R. & D. Bolinger (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association* 12. 58-69.
- Dauer, R.M. (1983). Stress-timing and Syllable-timing Reanalyzed. *Journal of Phonetics* 11. 51-62.
- Dauer, R.M. (1987). Phonetic and Phonological Components of Language Rhythm. In *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 5, 447-450. Tallinn: Academy of Sciences.
- Donovan, A. & Ch. J. Darwin (1979). The perceived rhythm of speech. In *Proceedings of the 9th International Congress of Phonetic Sciences*, vol. 2, 268-274. Copenhagen: Institute of Phonetics.

- Drake, C. & C. Palmer (1993). Accent structure in music performance. *Music Perception* 10(3). 343-378.
- Erickson, F. (1982). Money tree, lasagna bush, salt and pepper: Social construction of topical cohesion in a conversation among Italian Americans. In D. Tannen (ed.), *Analyzing discourse: Text and talk*, 43-70. Washington, D.C.: Georgetown University Press.
- Erickson, F. & J. Shultz (1982). *The Counselor as Gate Keeper: Social Interaction in Interviews*. New York: Academic Press.
- Faure, G., D.J. Hirst & M. Chafcouloff (1980). Rhythm in English: Isochronism, pitch and perceived stress. In L.R. Waugh, & C.H. van Schooneveld (eds.), *The Melody of Language*, 71-79. Baltimore: University Press Park.
- Fon, J. (2006). Cross-dialectal turn exchange rhythm in English interviews. In R. Hoffman, & H. Mixdorff (eds.), *Proceedings of the 3rd Speech Prosody Conference*, PS4-16-179. Dresden: TUDpress.
- Fusaroli, R., B. Bahrami, K. Olsen, A. Roepstorff, G. Rees, Ch. Frith & K. Tylén (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science* 23(8). 931-939.
- Fusaroli R., J. Rączaszek-Leonardi & K. Tylén (2014). Dialog as interpersonal synergy. *New Ideas in Psychology* 32. 147-157.
- Fusaroli R. & K. Tylén (2015). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*. 1–27.
- Garrod, S. & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27. 181-218.
- Grabe, E. & E.L. Low (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (eds.), *Papers in Laboratory Phonology* 7. The Hague: Mouton de Gruyter. http://www.phon.ox.ac.uk/~es-ther/ivyweb/Grabe_Low.doc.
- House, J. (2007). The role of prosody in constraining context selection: a procedural approach. *Nouveaux cahiers de linguistique française* 28. 369-383.
- Konvalinka, I., D. Xygalatas, J. Bulbulia, U. Schjødt, E.-M. Jegindø, S. Wallot, G. Van Orden & A. Roepstorff (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences* 108. 8514-8519.
- Kousidis, S. & D. Dorran (2009). *Monitoring convergence of temporal features in spontaneous dialogue speech*. Dublin: UCD Working Papers.
- Krauss, R.M. & S. Weinheimer (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology* 4. 343-346.
- Lea, W.A. (1974). *Prosodic Aids to Speech Recognition: IV. A General Strategy for Prosodically-Guided Speech Understanding*. Univac Report PX10791, St. Paul, Minnesota: Sperry Univac, DSD.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics* 5. 253-263.

- Levitan, R. & J. Hirschberg (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In R. Pieraccini, A. Colombo (eds.), *Proceedings of Interspeech 2011*, 3081-3084. Brisbane: International Speech Communications Association.
- Lloyd James, A. (1940). *Speech Signals in Telephony*. London: Pitman.
- Local, J. (2003). Phonetics and talk-in-interaction. In M. J. Solé, D. Recasens, & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 115-118. Barcelona: Casual.
- Low, E.L. & E. Grabe (1995). Prosodic patterns in Singapore English. In E. Kjell and P. Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 3, 636-639. Stockholm.
- Low, E.L., E. Grabe & F. Nolan (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language & Speech* 43(4). 377-401.
- McFarland, D.H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language and Hearing Research* 44. 128-143.
- Nespor, M. (1990). On the rhythm parameter in phonology. In I.M. Roca (ed.), *Logical issues in language acquisition*, 157-175. Dordrecht: Foris.
- Nespor, M. & I. Vogel (1986). *Prosodic phonology*. Dordrecht: Foris.
- O'Connor, J.D. (1965). *The perception of time intervals*. (Progress Report 2), London: Phonetics Laboratory, University College.
- Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustic Society of America* 119. 2382-2393.
- Pardo, J.S., R. Gibbons, A. Suppes & R.M. Krauss (2012). Phonetic convergence in college roommates. *Journal of Phonetics* 40. 190-197.
- Patel, A.D. & J.R. Daniele (2003). An empirical comparison of rhythm in language and music. *Cognition* 87(1). B35-B45.
- Pike, K.L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Port, R. (2003). Meter and speech. *Journal of Phonetics* 31. 599-611.
- Port, R., F. Cummins & M. Gasser (1996). A Dynamic Approach to Rhythm in Language: Toward a Temporal Phonology. In B. Luka, & B. Need (eds.), *Proceedings of the Chicago Linguistics Society*, 375-397. Chicago: Dept. of Linguistics, University of Chicago.
- Propp, V. (1928). *Morphology of the Folktale*. Austin: University of Texas Press, 1958.
- Ramus, F., M. Nespor & J. Mehler (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 72. 1-28.
- Reed, B.S. (2010). Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics* 42(4). 1037-1059.
- Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (ed.), *Linguistic controversies: essays in linguistic theory and practice in honour of F.R. Palmer*, 73-79. London: Edward Arnold.

- Russo, M. & W. Barry (2008). Isochrony reconsidered: Objectifying relations between rhythm measures and speech tempo. In P. A. Barbosa, S. Madureira & C. Reis (eds.), *Proceedings of the Speech Prosody 2008 Conference*, 419-422. Campinas.
- Sacks, H., E. Schegloff & G. Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50. 696-735.
- Shen, Y. & G.G. Peterson (1962). Isochronism in English. *University of Buffalo Studies in Linguistics - Occasional papers*, 9. 1-36.
- Turvey, M.T. (1977). Preliminaries to a theory of action with reference to vision. In R.E. Shaw & J. Bransford (eds.), *Perceiving, acting and knowing: Toward an ecological psychology*, 211-265. Hillsdale, NJ: Erlbaum.
- Uldall, E.T. (1971). Isochronous stresses in RP. In L.L. Hammerich, R. Jakobson & E. Zwirner (eds.), *Form and Substance*, 205-210. Copenhagen: Akademisk Forlag.
- Wenk, B.J. & F. Wiolland (1982). Is French really syllable-timed? *Journal of Phonetics* 10. 193-216.
- Wilson, M. & T.P. Wilson (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review* 12. 957-968.

Amedeo De Dominicis

University of Tuscia (Viterbo) and Centro Linceo "Beniamino Segre" (Roma)

Italy

email: dedomini@unitus.it