--- Accepted for publication in *Cognitive Computation*, 2014 ---

Title: Social Aspects of Entrainment in Spoken Interaction

Author: Štefan Beňuš
Affiliations: Constantine the Philosopher University, Nitra, Slovakia
Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
Address Stefanikova 67, 94974 Nitra, Slovakia
Contact: email: sbenus@ukf.sk, tel: + 421 37 64 08 455

Abstract

Speech entrainment is the tendency of interlocutors to become similar to each other during spoken interaction. Entrainment is a natural component of the cognitive system underlying communication, and the alignment of cognitive (para)linguistic representations between interlocutors is one way of conceptualizing it. Speech entrainment also plays an important social role, since humans perceive people who entrain to their speaking style as more socially attractive and likeable, more competent and intimate, and conversations with such partners as more successful. Furthermore, dis-entrainment might signal an increase in social distance and a negative attitude towards the interlocutor. Importantly for social robotics, humans also entrain to computer systems, and implementing this idea has brought improvements in several domains of human-machine interaction. This paper provides a targeted overview of advances in speech entrainment and argues that entrainment should be exploited in applications in which communication between humans and robots uses speech, as it opens up possibilities for developing and controlling social relations such as likeability and dominance and makes the applications more efficient.

Keywords: entrainment, speech, human-robot interaction, social cognition, dominance

1. Introduction

The goal of social robotics is to expand the applicability of robots in many domains of everyday life such as domestic care and work, healthcare, education, information exchange, communication, and companionship. Given this scope, many applications in *human-robot interaction* (HRI) target improvements in the social wellbeing of their users. The interactive nature of such applications requires a high degree of sophistication in the robot's model of the world and its functionalities for adapting to it. Since humans are the key component of the robot's environment, modelling of their social goals and developing a range of capabilities for jointly achieving them through HRI will be crucial for the success of such applications. Specifically, future social robots will have to be able to express and perceive the emotions and states of their interlocutors, exhibit distinctive personalities and characters, establish and maintain social relationships using human natural embodied systems such as speech, gaze, or gestures, and foster these social relations through acting together with humans. In short, social robots will have to have some form of *social cognition* in that they will need a pragmatic ability to act appropriately given certain situations and interlocutors [1]. These tasks pose enormous challenges for modelling robots' behaviour, given the vast degrees of freedom even in the extremely limited domains of HRI.

One of the ways in how this huge variability might be approached is to seriously examine and exploit the relationship between social cognition and *entrainment*. Entrainment is the tendency of interlocutors to become similar to each other in various aspects of their verbal and non-verbal behaviour. Modelling of social cognition in HRI requires understanding of communicative interaction [1], and one of the fundamental competences of this cognition is multi-modal behavioural entrainment [2]. In other words, any interaction of two cognitive systems, for example through speech, poses severe limits, and predictability, on what is said and done at any particular time. It is thus important to keep in mind that "truly random behaviour is rather rare, hence periodicity-within and synchronicity-between systems is, in practice, the norm" [3, p. 39]. Hence, interacting cognitive systems naturally synchronize, i.e. entrain to one another; this joint embodied acting together facilitates social cognition between the two systems and opens up possibilities for implementing natural constraints on the degrees of freedom when modelling their cognitive abilities.

The focus of this paper is on speech entrainment in which interlocutors converge on linguistic and paralinguistic aspects of speaking. In other words, speakers adapt their communicative behaviour to the behaviour of their conversational partner(s); see e.g. [4] for an extensive review of the literature. The goal of the current paper is to review selected advances and concepts in the area of speech entrainment and argue that the naturalness and effectiveness of applications in social robotics will be enhanced and social bonds between human and robot better controlled if they exploit speech entrainment. This goal contributes to one of the general aims of the cognitive computation research program characterized in [5]: "One feature in particular that must be developed is the ability in the machine to discern and empathise with the mental state of others with which it is in interaction, both machines and humans" [5: 4-5]. Moreover, entrainment might provide a useful concept for research into the cognitive architecture of HRI.

Rather than an extensive overview of the growing field of entrainment, this paper provides a targeted discussion focusing on speech, with the aim of stimulating inter-disciplinary exchange and cross-fertilization among the various disciplines involved in cognitive computational approaches to social robotics. The paper starts with a review of research on speech entrainment in human-human interaction (Section 2) followed by human-machine interaction (Section 3). Section 4 highlights the usefulness of speech entrainment for applications in social robotics and discusses in detail the notions of dominance and system efficiency. Section 5 discusses the role of entrainment in models of social cognition and mentions insights or testing platforms it can offer. Section 6 presents a general discussion and concludes the paper.

2. Entrainment in human-human spoken interaction

This section briefly reviews several studies reporting on entrainment in acoustic-prosodic features, structural linguistic features, as well as conversational patterns in turn-taking management and non-verbal behaviour during spoken interaction. Although the majority of studies analyse entrainment as it occurs in English, entrainment is a cross-linguistic phenomenon, since it has been observed in many unrelated languages and cultures (see references in [4]). The universal nature of human entrainment is also supported in [5]. This reference argues that interpersonal communication is based on utilizing various prosodic non-linguistic features such as pitch, energy, or duration for coordinating the behaviours of interlocutors. Speech entrainment can be considered a means for achieving such coordination. Importantly, because this coordination system is assumed to pre-date language evolutionarily, the coordination operates orthogonally from the semantic content and is thus assumed to be a universal cross-linguistic tendency.

2.1. Entrainment on acoustic and prosodic features

The acoustic signal, the primary source of information in spoken interaction, is extremely rich with patterns relating to temporal sequencing, intensity, and fundamental frequency. Many such acoustic features might convey linguistic as well as paralinguistic information at multiple levels. One of the main tasks concerning the broad spectrum of information that is provided by the acoustic signal is to understand the relationship between *redundancy* and *functional load* of these features. Consider, for example, vowel duration. This feature might convey lexical contrasts in languages such as Finnish or Slovak. But vowels in the stressed syllables of prominent words and syllables preceding prosodic boundaries are also longer than other vowels. Vowel duration thus participates also in identifying prominent words and structural units in the speech stream, which is essential for marking the informational structure of utterances necessary for the correct parsing of messages and thus decoding their meaning. Finally, vowel duration might also signal social attitudes such as boredom or engagement that are linked to variability in speech rate. Importantly, each of the abovementioned functions is signalled not only by vowel duration but also by other complementary acoustic features. This produces an intricate system of features that are functionally loaded, i.e. those participating in signalling multiple functions, and redundant, i.e. when one function is signalled by many features. Following [5] we assume that the core communication system between humans can be conceptualized as the coordination of behaviours among people, and the acoustic signal of spoken communication, specifically the redundancy of the acoustic and prosodic features mentioned above, provides ample affordance for interpersonal coordination.

In an influential paper, [7] observed that individual words spoken during mutual interaction become more similar in their phonetic and prosodic characteristics compared to the same words spoken before or after that interaction. The author concluded that talkers involved in collaborative spoken interaction increase similarity in their phonetic repertoires. Methodologically, [7] assessed entrainment by employing human *perceptual* judgments and testing similarity between word pairs. Entrainment has also been demonstrated through speech *production* by extracting and analysing features from the acoustic or articulatory records of speech, e.g. spectral features of vowels [8]. Similar observations were reported for accent and other socio-phonetic variables (e.g. [9]).

Considering now the scope of application, this non-linguistic entrainment during spoken interaction takes place both locally, i.e. at each turn-exchange between interlocutors, as well as globally, when entire conversations are examined. For example, [10] conceptualized entrainment as arising through similarity, convergence, or synchrony between speakers. They found significant degrees of entrainment, observable most clearly using features based on voice intensity but also in pitch, speaking rate, and voice quality, at both turn and conversation levels. They observed that the local entrainment at the turn-exchanges was more robust than when the unit of analysis was the entire conversation. Furthermore, [11] hypothesized that speech entrainment reflects social processes used to achieve strategic goals and that it increases over time within an interaction. They examined consistency in the direction of stylistic shifts, with "style" being conceptualized with a 70-dimensional vector of acoustic prosodic features extracted from the audio signal for each turn. Their approach was able to identify real vs. constructed pairs of interlocutors based solely on their style entrainment metric. Finally, a series of studies analysed entrainment on pitch and energy prosodic features in a corpus of married couples' problem-solving sessions [12, 13]. They used features extracted at the level of individual turns and sophisticated mathematical techniques, such as principal component analysis (PCA) [12] or combining correlation, mutual information, and coherence measures [13], for quantifying multi-dimensional speech entrainment. These studies exemplify a much needed approach in which rigid computational techniques are used for studying entrainment, and they can provide a "bridge between psychologists and engineers to help bring objective insight into human interaction" [13: 796].

The brief review above shows that speech entrainment in dyadic conversations takes place on the levels of individual words, prosodic features, overall speaking style, and locally at turn-exchanges as well as globally in entire conversations. The field has moved from documenting entrainment in rather small corpora of dedicated laboratory-elicited speech, often requiring manual labelling, to using sophisticated computational methods capable of extracting features directly from the audio signal of natural spontaneous conversations and assessing entrainment through a combination of features. These techniques thus offer possibilities for assessing entrainment at various granularities: locally at turn exchanges or globally per conversation, using features individually or as multi-dimensional vectors, or addressing particular aspects of speech features (e.g. intensity) vs. analysing more holistic aspects of 'style'. Given this varied nature of speech entrainment in human-human conversations, robust techniques for fast real-time implementation of entrainment at various granularities in human-machine dialogue systems is envisioned in the future.

Another area for which future progress is critical is adaptation to situational and environmental factors. Certain individual features, such as intensity, are heavily situation-dependent both when speech is produced and decoded. Situational factors affecting speech intensity include proximity of interlocutors, which commonly changes dynamically during spontaneous interactions, noise levels, degree of fatigue, and others. Yet, several studies observed speech intensity to provide best affordance for entrainment. Hence, successful utilization of speech entrainment increases demands for robust techniques for dealing with situation and environment adaptation.

2.2. Entrainment in other aspects of speaking

In addition to the acoustic and prosodic characteristics of speaking, entrainment has been observed in many linguistic features; see [14] for a recent review in this area. For illustration, [15] analysed *lexical entrainment*, which is the tendency of speakers to interactively converge on a name for a particular object and its subsequent conceptualization. They observed that once speakers establish such convergence, they re-use and simplify them and ultimately may abandon them for new mutually agreed names.

Entrainment takes place not only in content words and novel objects but has also been observed for function words such as pronouns, auxiliary verbs, or prepositions. For example, [16], working on transcripts from Supreme Court oral arguments, observed that Justices and lawyers entrain in terms of their patterns in the use of function words. It seems that these words express speakers' spatial and temporal points of view as well as interactional features, thus providing great affordance for interpersonal coordination and entrainment.

Entrainment has also been observed for syntactic structures or rules. For example, [17] conducted a laboratory experiment in which a subject and a confederate took turns in describing pictures; the confederate varied the form of di-transitive structures (e.g. ‘I gave him the apple’ or ‘I gave the apple to him’). The results showed that the syntactic structure of the confederate’s description significantly affected the syntactic structure of the speaker’s subsequent descriptions. In a more general study of syntactic entrainment, [18] analysed two large corpora of spoken interactions (the task-oriented Map Tasks and the more spontaneous Switchboard) with automatic syntactic parsing. They hypothesized that entrainment and mutual understanding are facilitated through *syntactic priming* in which an application of a particular syntactic rule or structure from one speaker triggers the use of the same rule/structure by the other speaker. Their data supported this hypothesis, since the probability of structural repetition between speakers decreased with the distance between the first and subsequent use of a particular syntactic structure. Interestingly, this effect was much stronger in task-oriented dialogues than in spontaneous ones. This is an important observation, since entrainment is high in situations that require a common situational model, which are precisely the situations at the core of the application of social robotics.

Another factor supporting the cognitive importance of entrainment comes from studies looking beyond traditional linguistic representations such as lexical words or syntactic structures, but which rather examine entrainment at a conceptual level. Lexical entrainment in [15] led to the creation of conceptual pacts between interlocutors, which were important for the successful completion of tasks. Similarly, [19] analysed ways in which interlocutors described their positions in a cooperative maze game. They found that partners entrained on their spatial perception of the maze, such as treating it as a coordination system or a route-system, which did not necessarily include using the same words or syntactic structures. These are examples of *conceptual entrainment*.

Finally, extending the scope of observation beyond words and sentences, interactional features relating to turn-taking management and common ground establishment are also subject to entrainment. For example, [20] looked at the timing of turn-initial single-word utterances. The target words included affirmative cue words such as *okay* or *mhm* and conversation fillers such as *uh* and *um*. They argued that entrainment, taken as the temporal and rhythmical incorporation of these turn-initiations with respect to the rhythm of the preceding turn, participates in online negotiation of dominance relationships linked to floor-control, as well as in dynamic creation of mutual common ground. Furthermore, people were found to converge on, and synchronize in, turn-latencies when evaluated both with static measures across entire conversations [21], and as dynamic online synchronizing [22]. Moreover, speakers were also found to entrain the pitch of backchannels to speech at the end of the preceding turn [23].

To sum up, entrainment is common at linguistic levels such as lexical choice or the use of syntactic structures, semantic and conceptual representations, as well as in broader pragmatic domains such as turn-taking management and common ground establishment. The observation that entrainment is stronger in task-oriented dialogues than in spontaneous ones is important for HRI, since entrainment is high in situations that require a common situational model, which are precisely the situations at the core of the application of social robotics.

2.3. Entrainment in other modalities

Although this paper focuses on speech, entrainment is pervasive also in other modalities and behaviours. For example, the often cited study [24] reported that people often mimic the postures, mannerisms, and facial expressions of their interaction partners and called it the ‘chameleon effect’. Examples included foot wagging, and touching hair or face. Interestingly, the setting for all experiments included spoken dyadic interactions. It is thus plausible that the speech of the participants displayed evidence of entrainment as well, and also that speech serves as the basic signal for entrainment in other modalities.

This hypothesis is supported by the findings of [25]. They investigated the postural movements of people involved in a cooperative verbal task and found that people who talk to each other develop interpersonal coordination in postural sway. Importantly, this coordination developed irrespective of the body orientation of the speakers, i.e. whether they faced each other or not. Hence, even in the absence of the visual modality, speech served as the coordination medium for behavioural entrainment. Moreover, it is well known that precise coordination of verbal and non-verbal actions (such as head nods or gaze shifts) is crucial for successful interactional communication; see for example [26] who showed this for human-human as well as human-robot interactions in the domain of museum and exhibition guides, or [27] as a recent study of coordination between co-speech gestures and prosodic events. Finally, married couples who have lived with each other for 25 years or more tend to develop similar facial features [28]. The authors suggest that the similarity in facial features stems from long-term empathizing with each other. Importantly for this paper, this empathizing is assumed to be accompanied by the copying of each other's facial expressions. Hence, entrainment in certain patterns of smiling or frowning over long periods of time creates similar patterns of wrinkles on the face.

Although the studies reviewed in this subsection are consistent with the hypothesis that speech drives interactional entrainment, entrainment also takes place in situations when no speech is involved. For example, days-old infants entrain in crying [29] or in facial gestures such as smiling or sticking out tongues [30]. People also commonly entrain in various kinds of bodily movement activities such as sport or dance, possibly through their ability to entrain to external rhythmical or oscillatory stimulus [31]. Additionally, when speech and even visual cues such as facial expressions and twitches are removed from the recordings of dyadic interactions, untrained judges still perceive interpersonal entrainment in a virtually identical way to the entrainment perceived in unmodified full recordings [32]. Hence, it seems that entrainment takes place during communicative interactions and speech is a natural, albeit not necessary, means for producing and perceiving entrainment. More experimental future research investigating the effect of various modalities individually, and in combination, on the degree and kind of entrainment is needed to pave the way for the building of adaptive multi-modal applications utilizing entrainment.

3. Entrainment in human-machine interaction

A possible way of approaching entrainment in human-machine interaction employs directionality. In a simplified uni-directional view, humans might entrain to machines, or machines might entrain to humans.[1] Evidence for both can be found. In the first case, when users entrain to dialogue systems, one of the features showing the beneficial effects of entrainment is speech rate. If the speech generation engine of a dialogue system produces speech whose rate is close to the optimal rate for the system's speech recognition engine, users tend to entrain by adapting their speech rate, which in turn results in improvements in speech recognition [33]. Interestingly, speakers are not consciously aware of their adjustments, and the human-machine interaction thus maintains naturalness of dialogue. This study shows a path for the future development of human-computer interfaces: "Users could thus be subtly influenced to adapt their speech to better match the current capabilities of the system, so that errors can be reduced and the overall quality of the human–computer interaction is improved" [33, p.2453].

In addition to speech rate, many other features show the promising effects of human entrainment towards a system. For instance, when interacting with animated characters, 7–10-year-old

---

[1] More natural dynamic and bi-directional aspects of entrainment in HRI will be briefly discussed in Section 6.

children displayed robust entrainment to the intensity of speech generated by the character [34]. Moreover, this entrainment in loudness as well as in response latencies improved the experience the children had from interacting with the animated characters. Regarding entrainment in lexical choice, [35] showed that system questions influence the lexical choices of user answers. Finally, [36] showed that entrainment on prosodic features positively correlates with learning gain when students interact with an automatic tutor.

The idea of shaping human behaviour through mutual entrainment has also been employed to improve human-robot interactions. For example, in instructions to move objects, entrainment in speech on lexical choices, or in vision on pointing or gaze gestures, led to smoother human-robot communication [37].

In the second direction, when machines entrain to humans, entrainment in phrasal and lexical choices was shown to be critical in situations when the conversation between the system and the human might break up [38]. Additionally, when a spoken tutoring system entrained to its users, thus increasing the cohesion of the interaction, learning gains surpassed those of un-entrained student-tutor interactions [39]. Finally, [40] showed that entrainment of the system towards the lexical and syntactic choices of the user has a positive impact on the performance of spoken dialogue systems. Specifically, it improves the recognition of task-related concepts.

In the turn-management domain, the difficulty that current state-of-the-art human-machine spoken dialogue systems need to address is how to achieve well-coordinated interactions. A possible explanation for common unsatisfactory user experience lies in coordination problems in the exchange of speaking turns between system and user. For example, currently the most common method for determining when the user has yielded the speaking turn consists in waiting for a long pause. However, this strategy is rarely used by humans, who rely instead on other types of cues, including syntactic, prosodic and acoustic ones, to anticipate turn transitions [41]. If such cues could be modelled and incorporated into future social robots, it would be possible to make faster and more accurate turn-taking decisions, thus making interactions more fluent [42], and consequently facilitate other kinds of prosodic entrainment on pitch or rhythmical and metrical features of speech.

Several studies also compare entrainment of humans when interacting with computers and other humans. In the area of lexical entrainment, [14, 43] showed that humans entrain more to computers than to other humans, which they analysed as a human’s effort to enhance communicative success. Importantly for social robotics applications, the degree of human entrainment was negatively correlated with the beliefs of the human regarding the communicative competence of the interacting cognitive agent: subjects entrained the most when they believed they were interacting with a computer with low communicative capacity; they entrained the least when believing they were interacting with a human; and an intermediate degree of entrainment was observed for subjects who believed they were interacting with computers with high communicative capacity [43]. This result, together with the observation that the degree of entrainment on linguistic structures tends to be greater in task-oriented scenarios than in spontaneous dialogues strengthens the argument for the plasticity of human cognitive capacities related to entrainment and thus the suitability of entrainment utilization for social cognition applications.

Finally, research on multi-modal entrainment in speech and gesture between humans and embodied communicative agents (ECA) represents an important step toward designing robots with entrainment capabilities enhancing their social cognition. Reference [44] provides an overview of the work in this area and describes an approach in which bottom-up perception of co-speech gestures combines with flexible top-down production of speech and gestures into an embodied coordination model in which the human and the ECA dynamically entrain their behaviour, and consequently beliefs and attitudes, through shared embodied sensorimotor structures.

To summarize, the ubiquity of entrainment in human-human dialogues transfers also to human-machine spoken interactions. Both directions of entrainment take place: humans entraining to machine and vice-versa. Moreover, entrainment facilitates the success and efficiency of human-machine interactional systems and opens opportunities for exploring embodied social cognition as the core underlying principle of human-machine interaction.

4. Entrainment is relevant for social robotics

4.1. Speech entrainment affects social relationships

The link between entrainment and social relationships has been proposed and convincingly developed over the last two decades by H. Giles and colleagues in their Communication Accommodation Theory (CAT) [45, 46]. The core idea behind CAT is that the degree of entrainment (accommodation) the speaker exhibits toward an interlocutor may be used as one of the means for achieving a desired social distance between the speaker and the interlocutor. In general, more entrainment leads to smaller social distance and thus a more positive outcome of the interaction. Hence, (speech) entrainment reflects speakers' need for social integration or identification with another.

The link between the degree of entrainment and positive social outcome has been suggested in several studies. For example, interlocutors whose behaviour displays a greater degree of entrainment have been found to conduct more successful and natural dialogues [47, 18, 48]. For specific domains of entrainment, subjects who entrain on speech rate are perceived as more socially attractive [49], viewed as more competent [50], and also more intelligent and supportive [45]. Entrainment also leads subjects to like their conversational partners more and have smoother interactions [24].

In studies of married couples' problem-solving, already discussed in Section 2.1, entrainment on pitch and energy prosodic features was found to positively correlate with high positive attitude [12], and entraining spouses were rated as having positive emotion [13], compared with low entrainment correlating with negative attitude and negative emotions. In the domain of child-robot interaction, entrainment facilitated increased engagement in social interaction [51]. Finally, [52] studied the relationship between entrainment on acoustic and prosodic features and independent annotation of speakers in four areas: their likability, dominance, encouraging behaviour, and awkwardness. They found that greater entrainment, especially on intensity and pitch, correlated with more likable and encouraging personalities. This applied mostly to female-male interactions and less so to same-gender interactions.

There is also some evidence that speakers are in full control of entrainment and may use it not only to decrease social distance but also consciously dis-entrain to show dislike and distance themselves from the interlocutor. For example, Welsh subjects broadened their Welsh accent significantly when interviewed by an arrogant interviewer with a strong English accent who called Welsh "a dying language with a dismal future" [53].

Finally, degree of entrainment correlates not only with social distance, belief about communicative competence, or likeable and engaging personalities, but it might also correlate with empathy; [28] discussed above.

The results of studies briefly reviewed in this section support the basic premise of CAT and indicate that there is a lawful positive relationship between the degree of human-human entrainment during spoken interactions on the one hand and decreased social distance, conceptualized as positive social traits such as perceived competence, supportiveness, likeability, encouraging nature, positive polarity, and others, on the other hand. What is needed in the future is more rigid experimental testing of the hypothesis, for example employing the Wizard of Oz paradigm, that the manipulation of the entrainment degree in human-machine interactions yields a systematic relationship to the core aspects of social cognition such as empathy, trust, belief, communicative competence, and others.

4.2. Speech entrainment and dominance

After reviewing the role of entrainment in building and maintaining social relationships, this section explores in more depth the relationship between entrainment and dominance. This area has not received much attention in entrainment research, but it also provides opportunities for better understanding of social cognition in human-human communication and extending cognitive capacities of future robots deployed in the social wellbeing domain.

Dominance is a powerful social dimension. Reference [3] discusses differences between living and non-living organisms and observes that while non-living systems (e.g. clocks on the wall or robots) display mutual entrainment, in living social systems (e.g. humans) coupling among agents can be a one-way process resulting in group behaviour. The author further says that “Organisms literally manipulate each other in order to get them to do what they want – a form of entrainment that can be interpreted socially in terms of the establishment and maintenance of dominance relations.” In human-robot interactions, this difference between living and non-living organisms should be taken seriously, and the role dominance plays in robots’ cognitive capacities should be explored. Do we prefer docile companions that do as ordered and can empathize when we discuss our personal problems with them? In that case, and in a greatly simplified way, we subscribe to a model of social communication in which a human is dominant and a robot is submissive. However, one can also imagine situations in which we might want social robots to be assertive and authoritative – for example, if a robot assists the elderly in the proper taking of medications, or if a robot participates in crises such as rescue missions after natural disasters and must persuade people to follow its instructions. In these situations, we would want the robot to be dominant and elicit submissive behaviour from humans. Entrainment might be one of the handles for achieving such domain specificity in HRI.

Below I briefly review selected research on dominance as a communicative strategy and the relationship between entrainment and dominance. In line with the overall argument of the paper, I suggest that the exploitation of entrainment during spoken human-robot interaction, for example through dynamic interactional models such as those in [2, 44, or 54], will facilitate robot behaviour that is natural and adaptable to various social needs.

Let us consider first the interactional nature of dominance in interpersonal communication. References [55, 56] construe dominance as a dynamic multidimensional communicative act through which an interlocutor exerts power or influence over one or more conversational partners by displaying linguistic signals of dominance. The notions of multi-dimensionality and dynamic nature are central in this approach to dominance and warrant a brief discussion. Dynamic nature refers to the fact that spoken interactions evolve over time and so do negotiations of dominance. This allows the differentiation of dominance from more static notions, e.g. power status, that certainly affect the communicative strategies associated with dominance, but dominance is not necessarily dependent on them. For example, two speakers may start a dialogue with roughly equal power positions and finish with very different ones. In this sense, communicative markers of dominance are constantly present in spoken interactions and are subject to an ongoing process of negotiation.

The multi-dimensional nature of dominance refers to the multi-dimensionality of information encoded in the acoustic signal and the ability of humans to control these dimensions separately. Hence, one speaker may be more dominant in speech intensity, while the roles might be reversed in other dimensions such as turn-management. Finally, dominance as a communicative strategy is always construed in relational rather than absolute terms, whereas asymmetries in power or status, as static notions, can be determined from factors external to the dialogue. As a communicative act, dominance of interlocutors is only defined in relation to dominance of their conversational partners.

This view, in which dominance is as an asymmetrical relationship dynamically created or maintained through communicative interaction, complements dyadic power theory [57, 58], which sees dominance as a combination of personal and contextual characteristics. The personal characteristics refer to the constant individual features and personal traits that are independent of the situation with which the individual is faced. The contextual characteristics include the dominance or submissiveness of the individual's partner in the interaction and are elaborated with the proposals of [55, 56] reviewed above.

Consider now the relationship between dominance and speech entrainment. If entrainment is observed, it may suggest an asymmetrical relationship. More specifically, and following CAT, the degree of speech entrainment is hypothesized to negatively correlate with dominance: if a speaker adjusts more than his/her interlocutor, that speaker is likely to be perceived as less dominant than his/her interlocutor. In other words, the communicative dominance of a speaker may be achieved, in part, by manipulating the degree of entrainment to the linguistic behaviour of the communicative partner. Several experimental studies support this hypothesis. For example, in [20] we studied dominance in relation to the temporal and metrical aspects of floor-control in dyadic task-oriented spoken interactions without visual contact. We identified various strategies for turn-initiations, such as hurrying the speaker or incorporating turn-initial utterance with the rhythm of the preceding utterance, and analysed the degree of entrainment of the interlocutors with respect to their dominance independently assessed with communicative and relational measures. The results of both qualitative and quantitative analyses of floor-control suggested that more dominant speakers entrain less, and more submissive entrain more. Interestingly, more traditional measures of dominance in turn-taking behaviour, such as the overall number of turns, the frequency of interruptions, or the amount of speaking time, did not clearly differentiate the dominant and submissive speakers in that corpus.

The link between entrainment and dominance has also been observed in other types of interaction. Reference [59] examined the relationship between acoustic entrainment and independently assessed social status of the interlocutor in a corpus of 25 dyadic interactions between a talk show host and his guests. The measure of entrainment was based on non-verbal vocal features present in so-called long-term average spectra (LTAS) and associated mostly with pitch and energy. They reported that lower status partners entrained their voices to higher status partners.

In the judicial domain, and more specifically Supreme Court oral arguments, [16] analysed the degree of entrainment in the use of function words. They showed that the degree of entrainment matched the power relationship between Justices and lawyers: lawyers matched their style to Justices more than Justices did to lawyers, mirroring the fact that Justices have a higher status and a more powerful role in this communicative situation. In the same domain of Supreme Court oral arguments, [60] reported a weak, but statistically significant, correlation effect between localized entrainment on the quality of conversational fillers between lawyers and Justices on the one hand, and chances for the favourable vote of the Justice in that lawyer's case.

A reviewer points out that the co-occurrence of high dominance and low entrainment reported above might be dependent on situational context. For example, entraining to an interlocutor might be seen as a sign of dominance in teachers or debaters. This is a plausible proposal that merits experimental investigation. A dominant speaker might "exert power or influence over a speaker" in the definition above by employing entrainment to lessen the social distance to the interlocutor, which will subsequently produce more empathy, liking, engagement, and other positive social feelings of the interlocutor toward the dominant speaker. This social relation then offers favourable grounds for the dominant speaker to influence his/her interlocutor.

Limited available research on the entrainment-dominance relationship in human-machine interactions also suggests that the asymmetrical power relationship might not be so crucial. Reference [61] reported a study in which subjects, divided into dominant and submissive groups based on personality tests, were randomly paired with a dominant or submissive computer, implemented with a simple variation of assertiveness of language and a number of commands and turn-initiations. Subjects preferred similar computers of matched personality, and they were also more satisfied with those interactions compared to interactions with a mismatched computer. Hence more research is needed to understand factors affecting dominance as asymmetry between speakers or dominance construed as similarity-attraction.

Research reviewed in this section supports the general argument of this paper that the relationship between speech entrainment and social aspects such as dominance offers fertile ground for applications in social robotics. This is because the degree of entrainment, when conceptualized as a control parameter, might provide functionality for a robot's dialogue systems that is similar to the ubiquity, naturalness and effectiveness of the entrainment-dominance link in human-human conversations. Once able to control the entrainment degree, social robotics applications might also be deployed in situations requiring domain specificity in HRIs.

4.3. Efficiency

The final dimension to be discussed here that shows the relevance of speech entrainment for social robotics is system efficiency. Given the huge number of degrees of freedom in modelling the behaviour of automatic systems, meaningful and independently motivated constraints limiting this number are highly valued. If such constraints improve the naturalness of the system's behaviour in addition to improving its efficiency, they are then clear candidates for engineering applications. I suggest that in interactive spoken systems, whether embedded in robots or other automated spoken dialogues systems, entrainment provides such a constraint. As also suggested by [3], entrainment between the system and the human decreases the number of control variables and thus facilitates the establishment of a control subspace with fewer control variables. This is also helpful in an effort to lower the energy consumption of the system, since fewer variables translate to a lower information rate and subsequent energy saving. Viewed from a different perspective, entrainment between the human and the system increases predictability. Since fewer resources are needed to convey predictable information, the saved resources then might be used for transferring less predictable information. For the same cost, an entrained human-robot system would be more efficient and natural compared to an un-entrained one.

5. Cognitive and architectural aspects

The discussion so far has mentioned several aspects of speech entrainment related to social cognition. This section expands on this link by briefly exploring entrainment in relation to the cognitive architecture and modelling of HRI. The thrust of the argument is that entrainment fits with, and can offer novel insights within, the view of social cognition as emerging from associative and distributed processes in the memory and emotion substrates.

Speech entrainment offers a unique domain for the design and evaluation of the computational architecture underlying interactions between humans and cognitive social robots. One of the most influential theoretical models of linguistic entrainment in human-human dialogues [47] posits that entrainment takes place through (para)linguistic representations. In this sense, entrainment would fit into the more traditional approaches to cognitive (linguistic) systems, artificial intelligence, and human-computer interaction that attempt to find the best structural representation of the data and model the interaction with environment with discrete-like atemporal algorithmical changes over these representations; see [62], and also the problem of grounding in symbolic representations [63]. Given recent critique of traditional, "good old-fashioned" representational approaches to AI or HRI – e.g.[64] calls for hybrid approaches allowing representations within embodied and thus biologically grounded views of cognition reviewed also in [65] – and a natural fit between embodiment and entrainment, there are at least two approaches to modelling entrainment that do not require symbolic representations and algorithmic deterministic procedures.

The first one is described in [66]. It sees (social) cognition as a set of constraints on action that arise from low-level physical, and thus embodied, coordination that is formalized through the coupling of oscillators within the dynamic systems theory. For example, the observed entrainment in postural sway movements during spoken interactions [25] that was discussed in Section 2.3, can be seen as a mutual dynamical and embodied interaction between shared knowledge across the two speakers and respective swaying motions of the speakers.

The second way is based on the interaction of sub-symbolic components in the connectionist fashion. Reference [67] summarizes an approach that sees cognition through "a future architecture that is fundamentally sub-symbolic throughout, but which carries out cognitive processes we now envision as symbolic as emergent consequences of the sub-symbolic computations" [67, p. 20]. See also this general approach applied to the cognitive ability underlying speaking with the use of the dynamic systems theory in [68, 69], who argue that dynamics offers a uniquely suited formal language for maintaining coarse-grained, discrete-like representational categories that are, however, embedded within the continuous substance of the perception-production loop.

This general line of research is also in line with [2], who have already shown as a proof of the concept that sensory information derived from human-robot interactions can be used in such a way that the behaviour of the robot could be modulated to match the behaviour of the human. In this approach, cognition and memory have "a fundamentally associative structure that is the substrate for activation dynamics, rather than being a symbol processing system at its core" [2:31].

In addition to the role of memory and embodiment for cognition, [65] also argue that cognition is inseparable from emotion in a broader sense. Specifically, they link cognition to both "external" expressing and recognizing of emotions used primarily in social communication, and the "internal" emotion proper of "having emotions". Entrainment fits nicely also with this approach to cognition. This is because it participates in the external system of emotional cognition. Moreover, it is possible to think of entrainment as an added feature, in principle encapsulated and separable from other layers. This might be a useful view from an engineering perspective. However, as soon as a robot produces speech (or any social behaviour in general), a degree of entrainment is by definition present, even if no entrainment has been programmed into the robot. Hence, just like emotion should not be considered as mere colouring of an agent but an inseparable part of an agent's cognition [65, 70], speech entrainment is an integral feature of any dialogue between two cognitive entities.

In sum, entrainment is a natural capacity within social cognition and lends itself straightforwardly to the architectural designs of HRI of several theoretical approaches to cognition. Entrainment might also provide a unique testing ground for exploring the applicability of these approaches to modelling social cognition.

6. Discussion and conclusion

The paper presented an overview of research related to entrainment in spoken human-human and human-system interactions focusing on the link between entrainment and social cognition. The discussion underscored the approach to spoken language as a form of joint action that requires interlocutors to coordinate with each other in order to be successful [71]. Speech entrainment was conceptualized as a form of such coordination, and the discussion presented ample evidence that 1) entrainment is everywhere where spoken interactions take place both among humans as well as between humans and machines, 2) that it affects conversations at linguistic and paralinguistic levels both locally and globally, 3) that it can have an effect on various features of speech at various granularities, and 4) that the degree of entrainment is lawfully linked to social distance and the relationship between interlocutors. This evidence supported the overall argument of the paper that entrainment is a natural capacity of social cognition and should be exploited by applications in social robotics because it has the potential to improve their naturalness and effectiveness.

Another goal of the paper was to facilitate cross-fertilization among the fields, since engineers and roboticists might benefit from the recent advances of linguists and communication scientists. Moreover, the very concept of entrainment adds to some fragmentation, since closely related terms such as *Coordination, Adaptation, Accommodation, Alignment, Priming, Convergence, Congruence, 'the Chameleon Effect', Mimicry, Coupling, (Interactional) Synchrony, Mirroring,* and others, have been used in various disciplines and frameworks.

The paper also highlighted several core aspects of the link between speech entrainment and social cognition: cross-modality, dynamic nature, and adaptability. Regarding cross-modality, the present contribution focused mostly on entrainment in the spoken modality. Nevertheless, it also mentioned several links to entrainment in gestures, body posture, and other aspects of the visual modality, and some lawful ways these types of entrainment might be linked to speech entrainment. Given the benefits of cross-modal entrainment in real HRI applications in the social wellbeing domain [e.g. 72], cross-modal entrainment including the spoken modality offers a promising avenue for future developments in social robotics. One of the questions that will have to be addressed is how to weigh the degrees of entrainment of multiple modalities for different social goals in future HRI applications.

The second aspect of entrainment re-occurring throughout the discussion was its dynamic nature. This can be conceptualized as the importance of temporal and embodied views of interactions in contrast to static views. This difference can be seen both at the level of conceptual cognitive modelling and at the level of dynamic development of social relations. The former includes discussions over abstract static vs. embodied dynamic representations such as articulatory gestures, and recent advances in modelling interactions through dynamic oscillatory systems. The latter includes, for instance, dynamic development of dominance over time through the interaction of various factors.

The dynamic control of entrainment in HRI systems might alleviate one of the concerns regarding the nature of HRI. Several authors argue that robots should not look like, be treated like, or be expected to behave like humans [e.g. 73]. However, treating robots as social agents is in principle orthogonal to considering them to be human-like. A way of approaching this issue through entrainment control would be to first provide means for humans to entrain to robots if they are inclined to, showing thus the development of a positive social relationship, and only once the robot detects entrainment from the human (on a particular feature), the robot might deploy its entrainment towards the human. This bi-directional and dynamic nature of entrainment is typical for the cognitive system underlying human-human interactions; see for example the visual representation of dyadic entrainment over time in [22].

A successful deployment of embodied and dynamically modelled HRI applications is, however, also contingent on continued progress in understanding the constraints embodiment poses on entrainment and subsequent social bonds. For example, an artificial cognitive system might naturally speak more slowly due to increased cognitive load but maybe not due to being tired. Alternatively, the pitch patterns of human speech and their link to emotions are naturally linked to their biological substance [74, 75], but this link is not straightforwardly extendable to robots. Therefore, better understanding of human perception of entrainment in HRI and subsequent development of trust, dominance, and other social constructs, as constrained by the different nature of embodiment between human and robot, is needed.

The final aspect highlighted in the discussion concerns adaptability. Developing functionalities for achieving dynamic and adaptable entrainment in HRI applications represents both a promising as well as challenging area for future research. This conceptualization of speech entrainment in the design of HRI applications also allows for the deployment of one robot in several situations with different social goals. For example, as a medical companion to patients, high degrees of speech entrainment might facilitate the building of closeness and trust of patients, but low degrees of entrainment might be useful for ensuring proper intake of medication or responding to medical emergencies.

In conclusion, interaction between robots and humans from a particular functional and an emotional perspective was discussed. The paper argued that spoken interactions between humans and robots should utilize control of speech entrainment at various levels, since it allows designers to facilitate the development of desirable social traits of humans towards robots, such as likeability, attractiveness, dominance, and others. In short, HRI systems will be more efficient, successful, natural, and pleasing if they are equipped with speech entrainment control.

Acknowledgments

This work results from the project implementation: Technology research for the management of business processes in heterogeneous distributed systems in real time with the support of multimodal communication, ITMS 26240220060 supported by the Research & Development Operational Programme funded by the ERDF. The author is indebted to Julia Hirschberg, Agustin Gravano, and Rivka Levitan. All mistakes are the sole responsibility of the author.

References

1. De Jaegher H, Di Paolo E, Gallagher S. Can social interaction constitute social cognition? Trends in Cognitive Sciences, 2010; 14: 441-447.
2. Baxter PE, de Greff J, Belpaeme T. Cognitive architecture for human–robot interaction: Towards behavioural alignment. Biologically Inspired Cognitive Architectures 2013; 6: 30-39.
3. Moore RK. Finding rhythm in speech: A response to Cummins. Empirical Musicology Review 2012; 7(1-2): 36-44.
4. Hirschberg J. Speaking more like you: Entrainment in conversational speech. Proceedings of Interspeech 2011; pp. 27-31.
5. Taylor JG. Cognitive computation. Cognitive Computation 2009; 1: 4-16.
6. Pentland A. To Signal Is Human. American Scientist 2010; 98: 204-210.
7. Pardo JS. On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America 2006; 119 (4): 2382–2393.
8. Babel M. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. Journal of Phonetics 2012; 40: 177-189.
9. Delvaux V, Soquett A. The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. Phonetica 2007; 64: 145–173.
10. Levitan R, Hirschberg J. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions, Proceedings of Interspeech 2011. pp. 3081–3084.
11. Jain M, McDonough J, Gweon G, Raj B, Rosé C. An unsupervised dynamic Bayesian network approach to measuring speech style accommodation. Proceedings of the 13th ECACL 2012. pp. 787-797.
12. Lee CC, Katsamanis A, Black M, Baucom B, Georgiou P, Narayanan S. An analysis of pca-based vocal entrainment measures in married couples' affective spoken interactions. Proceedings of Interspeech, 2011.
13. Lee CC, Black M, Katsamanis A, Lammer A, Baucom B, Christensen A, Georgiou P, Narayanan S. Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. Proceedings of Interspeech, 2010.
14. Branigan HP, Pickering MJ, Pearson J, McLean JF. Linguistic alignment between humans and computers. Journal of Pragmatics 2010; 42: 2355-2368.
15. Brennan SE, Clark HH. Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition 1996; 22(6):1482–1493.

16. Danescu-Niculescu-Mizil D, Lee L, Pang B, Kleinberg J. Echoes of power: Language effects and power differences in social interaction. *Proceedings of the 21st international conference on World Wide Web*; 2012. pp. 699-708.
17. Branigan HP, Pickering MJ, Cleland AA. Syntactic co-ordination in dialogue. Cognition 2000; 75:B13–25.
18. Reitter D, Keller F, Moore JD. Computational modelling of structural priming in dialogue," Proceedings of HLT/NAACL 2006; pp. 121-124.
19. Garrod S, Anderson A. Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. Cognition 1987; 27 (2): 181-218.
20. Beňuš Š, Gravano A, Hirschberg J. Pragmatic aspects of temporal entrainment in turn-taking. Journal of Pragmatics 2011; 43(12): 3001-3027.
21. ten Bosch L, Oostdijk N, Boves L. On temporal aspects of turn taking in conversational dialogues. Speech Communication 2005; 47: 80–86.
22. Edlund J, Heldner M, Hirschberg J. Pause and gap length in face-to-face interaction. In Proceedings of Interspeech, 2009; pp. 2779-2782.
23. Heldner M, Edlund J, Hirschberg J. Pitch similarity in the vicinity of backchannels. In Proceedings of Interspeech, 2010; pp. 3054-3057.
24. Chartrand T, Bargh J. The chameleon effect: The perception-behavior link and social interaction. Journal of Personality and Social Psychology 1999; 76:893-910.
25. Shockley K, Santana MV, Fowler, CA. Mutual interpersonal postural constraints are involved in cooperative conversation. Journal of Experimental Psychology: Human Perception & Performance 2003; 29: 326-332.
26. Yamazaki A, Yamazaki K, Burdelski M, Kuno Y, Fukushima M. Coordination of verbal and non-verbal actions in human-robot interaction at museums and exhibitions. Journal of Pragmatics 2010; 42(9): 2398-2414.
27. Loehr D.Temporal, structural, and pragmatic synchrony between intonation and gesture. Laboratory Phonology 2012; 3(1): 71-89.
28. Zajonc RB, Adelmann PK, Murphy ST, Niedenthal PM. Convergence in the physical appearance of spouses. Motivation and Emotion 1987; 11(4): 335-346.
29. Simmer ML. Newborn's response to the cry of another infant. Developmental Psychology 1971; 5: 136-150.
30. Melzoff AN, Moore MK. Imitation of facial and manual gestures by human neonates. Science 1977; 198: 75-78.
31. Phillips-Silver J, Aktipis A, Bryant G. The ecology of entrainment: Foundations of coordinated rhythmic movement. Music Perception 2010; 28(1): 3-14.
32. Bernieri FJ, Davis JM, Rosenthal R, Knee CR. Inter-actional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. Personality and Social Psychology Bulletin 1994; 20 (3): 303-311.
33. Bell L, Gustafson J, Heldner M. Prosodic adaptation in human-computer interaction. In Proceedings of International Congress of Phonetic Sciences 2003. pp. 2463-2466.
34. Oviatt SL, Darves C, Coulston R. Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. ACM Trans. Comput.-Hum. Interact. 2004; 11(3): 300-328.
35. Gustafson J, Larsson A, Carlson R, Hellman K. How do system questions influence lexical choices in user answers? Proceedings of Eurospeech 1997.
36. Thomason J, Nguyen HV, Litman DJ. Prosodic Entrainment and Tutoring Dialogue Success. In: Yacef K, editor. Artificial Intelligence in Education (LNCS 7926) Berlin: Springer; 2013. p. 750-753.

37. Iio T, Shiomi M, Shinozawa K, Miyashita T, Akimoto T, Hagita N. Lexical entrainment in human-robot interaction: Can robots entrain human vocabulary? In: Proceeding of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009.
38. Porzel R, Scheffler A, Malaka R. How Entrainment Increases Dialogical Effectiveness. In International Conference on Intelligent User Interfaces, Workshop on Effective Multimodal Dialogue Interfaces, 2006.
39. Litman D, Friedberg H, Forbes-Riley K. Prosodic Cues to Disengagement and Uncertainty in Physics Tutorial Dialogues. In: Proceedings of Interspeech, 2012.
40. Stoyanchev S, Stent A. Lexical and syntactic priming and their impact in deployed spoken dialogue systems. In: Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2009.
41. Gravano A, Hirschberg J. Turn-taking cues in task-oriented dialogue. Computer Speech and Language 2011; 25(3):601-634.
42. Raux A, Eskenazi M. Optimizing the turn-taking behavior of task-oriented spoken dialog systems. ACM Transactions on Speech and Language Processing 2012; 9:1.
43. Branigan HP, Pickering MJ, Pearson J, McLean JF, Brown A. The role of beliefs in lexical alignment:Evidence from dialogues with humans and computers. Cognition 2011; 121: 41-57.
44. Kopp S. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. Speech Communication 2010; 52(6): 587-597.
45. Giles H, Mulac A, Bradac JJ, Johnson P. Speech accommodation theory: The first decade and beyond. In M. McLaughlin (ed.): Communication Yearbook, 10. Newbury Park, CA: Sage; 1987. pp. 13-48, 1987.
46. Giles H, Coupland N, Coupland J. Accommodation theory: Communication, context, and consequence. In: Giles H. Coupland N. Coupland J, editors. Contexts of accommodation: Developments in applied sociolinguistics. Cambridge: Cambridge University Press; 1991. pp. 1–68.
47. Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences 2004; 27:169–226.
48. Nenkova A, Gravano A, Hirschberg J. High frequency word entrainment in spoken dialogue. Proceedings of ACL/HLT, 2008.
49. Putnam W, Street RL. The conception and perception of noncontent speech performance: Implications for speech accommodation theory. International Journal of the Sociology of Language 1984; 46:97-114.
50. Street RL Jr. Speech convergence and social evaluation in fact-finding interviews. Human Communication Research 1984; 11:139-169.
51. Belpaeme T, Baxter P, Read R, Wood R, Cuayahuitl H, Kiefer B., et al. Multimodal child-robot interaction: Building social bonds. Journal of Human-Robot Interaction 2012; 1: 33–53.
52. Levitan R, Gravano A., Wilson L, Benus S, Hirschberg J, Nenkova A. Acoustic-Prosodic Entrainment and Social Behavior. Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2012. pp. 11–19.
53. Bourhis RY, Giles H. The language of intergroup distinctiveness. In Giles, H, editor. Language, ethnicity and intergroup relations. New York: European Association of Experimental Social Psychology; 1977. pp. 119-135.
54. Ziemke T, Lowe R. On the role of emotion in embodied cognitive architectures: From organisms to robots. Cognitive Computation 2009; 1: 104-117.
55. Poggi I, D’Errico F. Dominance signals in debates. In A.A. Salah et al., editors. *HBU 2010, LNCS 6219*. Berlin: Springer; 2010. p. 163–174.

56. Poggi I, D'Errico F, Vincze L. Agreement and its multimodal communication in debates: A qualitative qnalysis. Cognitive Computation 2011; 3:466-479.
57. Dunbar NE, Burgoon JK. Perceptions of power and interactional dominance in interpersonal relationships. Journal of Social and Personal Relationships 2005; 22: 231–257.
58. Dunbar NE, Bippus AM, Young SL. Interpersonal dominance in relational conflict: a view from Dyadic Power Theory. Interpersona 2008; 2(1): 1–33.
59. Gregory SW, Webster S. A nonverbal signal in voices of interview partners effectively predicts communication entrainment and social status perceptions. Journal of Personality and Social Psychology 1996; 70: 1231–1240.
60. Beňuš Š, Levitan R., Hirschberg J. Entrainment in spontaneous speech: the case of filled pauses in Supreme Court hearings. Proceedings of the 3rd IEEE Conference on Cognitive Infocommunications, 2012; 793-797.
61. Nass C, Moon Y, Fogg BJ, Reeves B, Dryer DC. Can computer personalities be human personalities?" International Journal of Human-Computer Studies 1995; 43(2):223-239.
62. Kemp C, Tenenbaum JB. The discovery of structural form. Proc Natl Acad Sci. 2008; 105(31):10687–92.
63. Harnad S. The symbol grounding problem. *Physica D* 1990; 42: 335-346.
64. Muller V. Interaction and resistance: The recognition of intentions in new human-computer interaction. In: Esposito A et al. editors COST 2102 Int. Training School (LNCS 6456), Berlin: Springer; 2011. p. 1-7.
65. Ziemke T, Lowe R. On the role of emotion in embodied cognitive architectures: From organisms to robots. Cognitive Computation 2009; 1: 104-117.
66. Shockley K, Richardson D, Dale R. Conversation and Coordinative Structures. Topics in Cognitive Science 2009; 1: 305-319.
67. McClelland JL. Is a machine realization of truly human-like intelligence achievable? Cognitive Computation 2009; 1: 17-21.
68. Gafos A. Dynamics in grammar: Comments on Ladd and Ernestus & Baayen. In: Goldstein L, Whalen D, Best C, editors. Varieties of phonological competence (Laboratory phonology No. 8). Berlin: Mouton de Gruyter; 2006. p. 51-79.
69. Gafos A, Beňuš Š. Dynamics of phonological cognition. Cognitive Science 2007; 30(5): 905-943.
70. Varela FJ, Depraz N. At the source of time: Valence and the constitutional dynamics of affect. Journal of Consciousness Studies 2005;12(8-10): 61–81.
71. Clark HH. Using language. Cambridge: Cambridge University Press; 1996.
72. Belpaeme T. et al. Multimodal Child-Robot Interaction: Building Social Bonds. Journal of Human-Robot Interaction 2012; 1(2): 33–53.
73. Dautenhahn K. Methodology & Themes of Human-Robot Interaction: A Growing Research Field. International Journal of Advanced Robotic Systems 2007; 4(1): 103-108.
74. Gussenhoven C. Intonation and interpretation: Phonetics and phonology. In Bel B., Marlien I. (eds.) *Proceedings of Speech Prosody*, 2002, p. 47-57.
75. Hirschberg, J. The Pragmatics of Intonational Meaning. In Bel B., Marlien I. (eds.) *Proceedings of Speech Prosody*, 2002, p. 65-68.