Dynamic Field Theory and Embodied Communication

Yuliya Sandamirskaya and Gregor Schöner

Institut für Neuroinformatik,
Universitätstr. 150, Bochum, Germany
yulia.sandamirskaya@neuroinformatik.rub.de,
gregor.schoner@neuroinformatik.rub.de
http://www.neuroinformatik.rub.de

Abstract. Dynamical Field Theory is a neurally based approach to embodied and situated cognition, in which information is represented in continuous activation fields defined over metric spaces. The temporal evolution of activation patterns under the influence of inputs and neuronal interaction is described by a dynamical system, whose stable states, localized peaks of activation, are the units of representation. This approach has been successfully used to capture many elementary forms of cognition. Communication poses the new challenge of understanding how different modalities can be integrated in a continuously unfolding communicative process. In this chapter we give a brief introduction to Dynamical Field Theory in embodied cognition, and discuss extensions of its ideas to embodied communication. We sketch a highly simplified example of how sequence generation may occur in dynamical fields. We apply these concepts to a specific exemplary problem in embodied communication, turn taking, the temporal structure of which we capture in a simple model.

Keywords: Neural dynamics, embodied cognition, modelling, turn taking.

1 Embodied Cognition and Embodied Communication

Over the last decade or so, a reexamination of our understanding of cognition and cognitive processes has begun which emphasizes that cognition takes place in organisms who act in complex, structured environments [1–3]. Cognition, in this view, unfolds in real time, continuously linked to sensory information and continuously coupled to motor systems, which impact on the sensed world [4]. People bring to any cognitive act a history of prior experience both in the sense of the immediate behavioral and sensory context, in which the cognitive act takes place, as well as in the sense of the longer personal history of learning and development, on which cognition builds. The embodied perspective on cognition calls into doubt the postulate of universal representations of knowledge, on which cognition can operate by processing information. Instead, this perspective demands that an understanding of cognition be based on principles of neuronal function, in particular, on the temporally and spatially continuous evolution of neuronal representations, which may be updated at any time by sensory signals and which remain linked to motor surfaces at all times [5].

Communication involves, of course, a particularly high level of cognition. Are the insights from the embodied and situated perspective on cognition relevant to understanding communication? As this books argues in the most varied ways, communication is embodied in the obvious sense that natural communication occurs between people with physical and physiological bodies, and their cognitive abilities originate from processes in their nervous systems. Each individual brings his or her particular history of behavior, memories, intellectual capabilities to a communicative situation. Communication is situated in the similarly obvious sense that it is acted out in a specific, structured environment, in which the interaction between communicating people or agents occurs. For each actor, the actions of other actors shape the environment, in which communicative processes are embedded. Communication is multimodal, including verbal but also many non-verbal sensory and motor dimensions such as gesture, mimic, and bodily pose. Communication happens at different level of awareness.

The most saliently embodied aspect of communication may be turn-taking [6]. In conversation between two or more actors, almost all time is taken up by explicit verbal communication. Speaker and listeners switch roles very quickly. Typical durations of the silences between such switches are of the order of 100 to 300 ms, often faster than even a simple reaction time [7]. Such smooth transitions may be viewed as evidence for anticipatory planning [8]. They require time-continuous monitoring of the speaker's behavior, picking up potentially subtle, graded cues as to when the turn will be yielded. Such cues include changes of prosody, gestures, gaze shifts toward the listener, implicit verbal cues such as filler utterances, or explicit verbal cues such as asking a question.

We will use turn-taking as an exemplary problem in embodied communication, around which we will highlight how concepts from embodied cognition may be useful to understand embodied communication. Our mission is to examine conceptual convergence and interchange between these two areas from the perspective of Dynamic Field Theory (DFT), a particular mathematically explicit theoretical framework within which concepts of embodied cognition can be made precise [9, 5, 4]. DFT and its neuronal basis will be briefly reviewed in the next section.

Three aspects of DFT seem to us potentially relevant to understanding embodied communication. The first is *autonomy*. In cognition, autonomy refers to the fact that cognitive processes unfold continuously in time on the basis of the current and past behavior and influenced by current and past sensory information. Dynamical systems thinking emphasizes autonomy, which may be contrasted with the input-compute-output perspective on which the framework of information processing is based. That acts of communication such as a lecture, a group discussion, or a dialog, unfold in real time is obvious. Such acts are not appropriately described as series of input-output mappings. Instead, the state of the complete system containing communicating partners and the environment impacts on how the communicative process unfolds. For instance, the inner state of the participants (their arousal, mood, knowledge, willingness to contribute to the discussion or to express an opinion), the nature of the interaction (e.g., eye contact, gesture, prosody), the environmental conditions (e.g., noise level, people in the background), as well as the behavioral and cognitive context (e.g., nature of the relationship, recent experiences) all may modulate the multidimensional behavior of the participants. Turn-taking might be the most accessible signature of autonomy.

The second aspect is *gradedness*. In Dynamic Field Theory (DFT), the state of a neuronal system may be varying along continuous dimensions, which include an "intensity" dimension that measures the potential impact of each factor. In communication, the states of the participants and of the environment may vary in graded fashion as well. For instance, gestures may differ in their extent and expressiveness, voices in volume and timbre, the strength of emotions and their expression through various channels may vary. Those graded factors may influence the time course of communication by shifting, for instance, points of turn taking, and the level or tone of the response. Graded variations like these may potentially lead to qualitative change, switching the overall shape of a communicative act, say from neutral to hostile.

Such a switch would be an incident of loss of stability. *Stability* is the property of states in an autonomous dynamical system of resisting change. In a stable state, small changes in conditions lead to small changes in the state of a system. Given the amazing flexibility of nervous system, any state that persists long enough to be observable must have some degree of stability, so that it resists the myriad influences that push the system in other directions. Stability plays a crucial role in cognitive systems that are situated within time varying real-world environments. In fact, autonomy, the continuous evolution of a system in contact with such environments, cannot lead to macroscopically coherent behaviors if mechanisms of stability do not protect the current state of the system from the ensemble of perturbations coming from within a system as well as from its environment. Stability is also a prerequisite for the coherence of higher levels of behavior to be preserved under processes of development, learning and adaptation. In Dynamic Field Theory (DFT), stability is shown to emerge from the underlying neuronal dynamics [10].

Stability is clearly a relevant concept to understand how processes of embodied communication may unfold continuously in time under the influence of multiple graded variables. States with a higher degree of stability may persist of larger periods of time and resist competing influences. In turn taking, for instance, very stable states of the communication system may prevent a transition to a new state, while less stable or even unstable states may be both the prerequisite to change as well as the mechanism through which change comes about.

A challenge for understanding embodied communication under the constraints of autonomy, gradedness and stability is sequence generation. Much of cognition, but certainly all communication involves sequential changes of state, whose serial order is typically relevant. Generating sequential shifts of states require that the states are released from stability. On the other hand, a sequence as a whole must resist perturbations and thus have some sort of stability. Understanding sequence generation from a neuro-dynamical perspective is not a solved problem, although a number of efforts exist [11–16] After reviewing the main concepts of Dynamic Field Theory (DFT), we will therefore provide a sketch of how sequence generation can be conceived of within DFT.

We will then explore how theoretical concepts from Dynamical Systems thinking may impact on our understanding of embodied communication. This we will done by providing an explicit mathematical model of turn taking, with which we attempt to account for some of the qualities and one quantitative feature of the phenomenon. The model is, however, largely metaphorical in nature. It is meant to illustrate, how the time structure of communicative processes may emerge from time-continuous processes, how categorical change in continuous representations may emerge from multiple possible causes and how all kinds of contributions to the system's dynamics may matter, not only contributions that are specifically linked to a particular contents. The dynamical systems metaphor, we thus aim to illustrate, promotes thinking about underlying forces and regularities, from which the complex, multi-facetted patterns of communicative behavior may emerge.

2 Dynamic Field Theory (DFT)

The following is a brief survey over DFT, which has been reviewed more extensively elsewhere [9, 17, 4]. The mathematics of DFT come from the field of dynamical neuronal networks, pioneered by [18], [19] and [20] and currently the preferred route of many computational neuroscientists to function (e.g., [21], [22]). On this basis, DFT is an approach that emphasizes concepts that align closely with the needs of experimenters in human embodied cognition. In DFT, the basis concepts are inspired by principles of neuronal function in the central nervous system, but behavioral experiments provide the major constraints for both modelling and theoretical thinking.

At the core of DFT is the notion of continuous activation fields. While these have been historically derived as an approximate description of cortical neurophysiology [19], they arise in DFT our of more abstract arguments linked to an analysis of embodied cognition. This will be our first concern. Next we will discuss the dynamics of activation fields based on inputs and interactions, emphasizing peaks as units of representation. Finally, we will take the reader through three instabilities of the dynamics of activation fields which are critical to understanding how sequences can be generated from attractor states that turn unstable.

In order to represent metric information in terms of dynamical state variables, we need two dimensions (Fig. 1). One is the metric dimension along which information is specified. To model communication, metric dimensions that may play a role include the direction of an deictic gesture, a range of visual expressions, a range of prosodic speech patterns, etc. Seemingly categorical aspects of communication such as the contents of a verbal message may likewise be thought of as embedded in an underlying continuum, which reflects relationships of semantic similarity [23, 24]. In general, the multi-dimensional metric space spans the range of possible communicative intentions as a whole, comprising all aspects of mental state that impact on the communicative process. For now, we shall be visualizing dynamic fields over a single dimension, however.

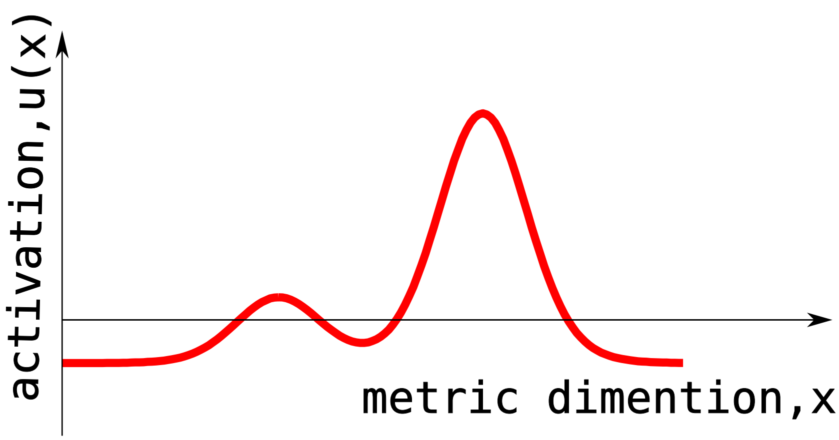

Fig. 1. A dynamical field is an activation pattern, $u(x,t)$, defined over metric dimension, x, at any moment in time, t. Peaks of positive activation represent a decision that the field has a well-defined state along the dimension, x, specified by the location of the peak in the field. Patterns of non-positive activation, by contrast, represent graded information typically derived from input.

The second dimension is the extent to which any given value along the metric dimension is currently active. This is the activation concept of cognitive science, known in this form also as the principle of space coding in neuroscience, according to which the location in the neural network determines what is encoded, while neuronal activation signals the absence or presence of information [25, 26]. The activation dimension may also represent graded values such as the strength of a particular dimension, confidence in an estimate, or how close a particular representation is to impact on the further evolution of a communicative process.

Activation fields evolve continuously in time under the influence of inputs and internal interactions. This evolution is described by a dynamical system. The mathematical description is motivated by the dynamic properties of neurons in the central nervous system [19, 20, 27, 10]. The fundamental property of the field dynamics, the stability of activation patterns, emerges from the biophysics of neurons irrespective of the specific neuronal model and its implementation details (see [10] for an argument).

In the absence of any inputs, the resting level of the dynamic field is therefore a stable state. Inputs may arise from sensory information or from other dynamic fields. Within the setting of communication, input is expected to reflect the impact of the other communicative partner, sensed in various ways. Additional inputs may reflect the influence of memory, knowledge, and the sensed environment. Inputs may be focused on particular field locations and thus specify a particular value of the metric dimension. Alternatively, inputs may be global (homogeneous), affecting the whole field and impacting on the dynamic regime in which the field operates.

Neuronal interaction is the dependence of the time course of activation at one field location on the current activation at other field locations. Neuronal interaction may stabilize localized peaks of activation, which are the stable objects that form the elementary units of representation in DFT. Nearby field locations are assumed to mutually excite each other, driving up activation, while distant locations are assumed to mutually inhibit, driving down activation. Only sufficiently activated locations contribute to interaction. This is modelled mathematically by applying a sigmoid nonlinearity to all activation variables contributing to interaction. This pattern of neuronal interaction is generic in the cerebral cortex, but can also be observed in many subcortical structures.

Two types of attractor solutions emerge from the interaction pattern in neural fields. Input-driven attractors are largely subthreshold patterns of activation in which the contribution of the neuronal interaction is negligible. In such states, the field merely filters external input, but does not make any selection or detection decisions. One may visualize these solutions as representing a passive, information processing mode, in which cognition is not yet engaged.

Self-stabilized attractors, by contrast, are localized patterns of activation with levels sufficient to engage interaction by exceeding the threshold of the sigmoid nonlinearity. Local excitatory interaction thus lifts activation within the peak beyond levels induced by input, while global inhibitory interaction suppresses levels elsewhere below the levels justified by the resting level or by inputs. Such peaks of activation are stable against small variations in local input as well as against weak competing inputs at other field locations. Such localized peaks of activation are the units of representation, their location encoding metric information about the underlying dimension while their level of activation indicates something like the strength, certainty, or intensity of the represented value.

In fact, self-stabilized peaks represent the outcome of various forms of decision making. This can be seen by noting that an instability connects the subthreshold, input-driven patterns of activation to self-stabilized peaks. The simplest form is the *detection*-instability, in which a localized external input is increased in strength (see [4] for a review of this and the following instabilities). At a critical input strength, the subthreshold pattern of activation becomes unstable and a localized peak forms. That peak remains stable, even if input strength

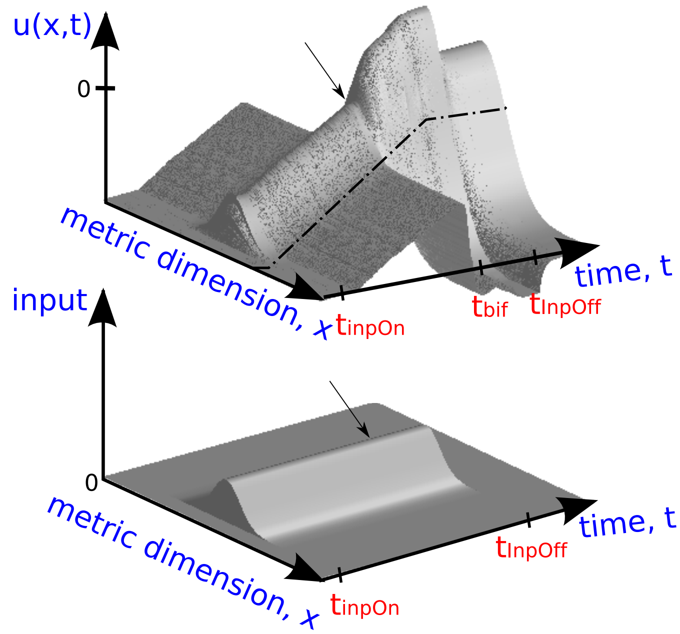

Fig. 2. A one-dimensional field, $u(x,t)$ evolving in time as described by the field dynamics. When localized external input arrives at time $t = t_{inpOn}$ (bottom plot), the field relaxes to the subthreshold pattern of activation defined by the input (top plot). At the same time the homogeneous resting level begins to be lifted gradually (the resting level of the field is marked with a dotted line). This induces an instability at time $t = t_{bif}$ (marked with an arrow on both plots) and the field relaxes to a new attractor, a self-stabilized peak dominated by interaction. Even later, at $t = t_{inpOff}$, when the external input is removed, the localized peak of activation is sustained by interaction.

drops again (within limits, see below). This provides a process model of the detection decisions central to most early psychophysics and often conceived of in terms of signal detection theory [4].

Another form of this instability is, at first, more surprising (Fig. 2). It arises starting from a subthreshold pattern of activation. If activation is now boosted by a homogeneous form of input (modelled, for example, by increasing the resting level of the field), the field location with highest subthreshold level of activation first pierces the threshold. This makes the subthreshold solution unstable, and a self-stabilized localized peak grows out of this event. Because the location of the peak is determined by preexisting inhomogeneities in the field, this type of instability could be viewed as a form of categorization [28, 4].

Another way of looking at this instability is that it amplifies small, subthreshold patterns of activation into macroscopic decisions, which can be acted out. Thus, if small traces of previous patterns of activation can be left by a simple learning mechanism, then this instability can activate such prior experience into units of representation. This opens the fields to generate long-term memory of their own behavioral history. In the context of communication such a mechanism can be used to model endogenous factors determining the course of an interchange, beyond a processing of only the incoming information.

If the general level of activation in the field is sufficiently high, interaction may enable the dynamic field to sustain a localized peak of activation even after the original localized input has been removed. Such sustained patterns of activation represent a form of metric working memory [5]. Thus information about past stimulation can be preserved over much longer time scales than the dynamic time scale of individual neurons.

Localized peaks may arise from yet another form of decision making, the *selection* among multiple localized inputs. Selection is also an elementary cognitive function that can be modelled with dynamic fields. This function too emerges from an instability. When two metrically close inputs are presented to the field, the detection instability will lead to a broader peak centered around an averaged location between the two peaks. When the metric distance between two localized inputs is larger, however, the dynamic field is bistable and has the potential to build a peak at either of the two locations. A single peak of activation, which emerges due to the detection instability, will be centered around one of the two locations specified by the inputs. The sites which pass the threshold and participate in interaction are too distant in this case in order to support each other due to the local excitation. Asymmetries in input, fluctuations, or prior activation history may favor one over the other states, but the far-reaching inhibition prevents simultaneous activation of both locations.

3 Sequence Generation within DFT

Generating sequences of states or actions is central to cognition. A train of thought could be viewed of as a sequence of mental states. Language is sequential at many levels — sequences of articulatory gestures, of sounds, of words or larger syntactical elements, of ideas or arguments. Goal-directed action is another important form of sequence generation within (embodied) cognition. Think of the actions needed to make coffee or to fetch an object from a cup-board [29]. In communication too we observe sequences of mental states, utterances, gestures, and other communicative acts.

Sequence generation is conceptually cheap in classical information processing approaches such as those based on the analogy with the digital computer. In fact, those approaches are essentially based on the concept of processing by moving from one step in a sequence to the next. This processing is, in a sense, atemporal, because it doesn't matter, how much real time elapses while such a step is made. The advance of information through the processing system itself marks time. In contrast, in an embodied vision of cognition, all processes are temporally autonomous, may at any time be linked to new sensory information or to ongoing motor behaviors or even to other processes running concurrently. Stability is critical for such highly interlinked processes to stay on track. Because stable states resist change, they are capable of persisting under time-varying conditions. But there is a dilemma. The very resistance to change makes it difficult to conceive of the generation of a sequence of states. After all, moving from one state to the next requires that the previous state be released from stability, that is, become unstable, so that the system is driven toward the next state.

That it is not impossible to reconcile this need for stability and instability in a dynamic approach to embodied cognition is illustrated by the simple model that will be briefly sketched now [30]. It is not in the nature of this chapter (and of the volume in which it appears) to provide a detailed mathematical description of this model (nor of next section's turn taking model). Therefore, we focus on sketching the main ideas of the model, explaining the overall architecture, variables, and dynamical principles. The equations are listed in the appendix in enough detail to be implementable by the mathematically skilled reader, but are probably not accessible to the typical reader.

The model consists of an ensemble of activation fields (Fig.3). All fields represent metric or categorical information relevant for what happens at any stage within a sequence of actions (e.g., feature values, movement parameters, descriptors of utterances or gestures). A set of fields (arranged in a stack in Figure 3) represents the ordinal position within the sequence. Each field in the stack is responsible for one step in the sequence. This encoding comes about through a coupling among the fields that guarantees their sequential activation: Each field provides spatially homogeneous, excitatory input to its "successor" field, that is, the field that represents the next ordinal position in the sequence. In addition, each field inhibits its predecessor. As a result, only one field in the ordinal stack can be activated at a time.

Beyond the stack of ordinal fields, one additional output or "motor" field represents the currently activated action plan (Fig. 3). This field receives localized (one-to-one) input from all ordinal fields. The motor field activates the sensori-motor action system that brings about the planned action. That sensori-motor system is modelled for now only by assuming that sensory feedback about the successful termination of an action is provided at a variable time after initiation of the action (a more specific model will be provided in the next section). This signal represents the "condition of satisfaction" [31] of the planned action and generates spatially homogenous excitatory input into all fields in the ordinal stack. This input triggers an instability through which the currently activated ordinal field becomes deactivated and its successor becomes activated. This is the elementary transition in sequence generation from step in a sequence to the next.

All activation fields are endowed with the neuronal interaction that leads to the selection of a single peak even when multiple localized inputs are present. As a result, except for the brief moment of transition from one step to the next, there is always a peak present in one field of the stack of ordinal fields and

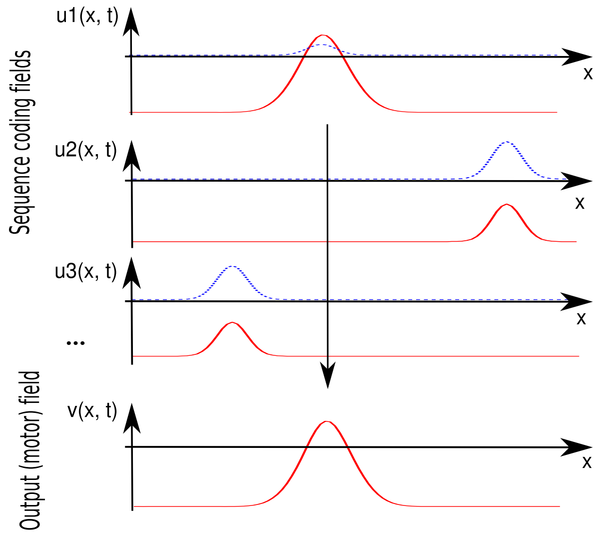

Fig. 3. A sequence generation model based on dynamic fields whose stable states are single localized peaks. A stack of such activation fields, $u_i(x,t)$ (solid lines), encodes the ordinal position of steps, $i = 1, 2, 3$, in the sequence. All fields are defined over a shared metrical dimension, x. The fields may be preactivated by localized input (dotted lines) representing information about what action is expected at each step in the sequence. Each field excites its successor and inhibits its predecessor, so that only a single field can be active at a time ($u_1(x,t)$ at the depicted instant). The output or "motor" field, $v(x,t)$ (bottom of stack) receives localized input from all ordinal fields and thus carries a peak matching the currently active ordinal field.

one matching peak in the motor field. Where the peaks are located within the fields and thus, what action is planned and triggered at any particular point in a sequence is determined by localized inputs into the fields. There are two potential sources of such localized input that determine the "contents" of the sequence: external input from sensory systems or internal input from a memory trace that represents earlier experience, when the sequence was first learned or imitated. In summary, localized input determines *what* is done at each step. Sensory feedback representing the "condition of satisfaction" determines *when* the transition to the next step is made.

The notion of a neural representation of the ordinal position of an item in a sequence is consistent with the neurophysiology of frontal cortex, in which neurons tuned to ordinal position have been found and lesions of which affect the capacity to generate sequential action [32, 33]. The proposed dynamical architecture addresses the fact that the "acting out" in time of action sequences is flexible, so that the system is capable of remembering where in a sequence it is even as variable amounts of time elapse while each individual action is realized.

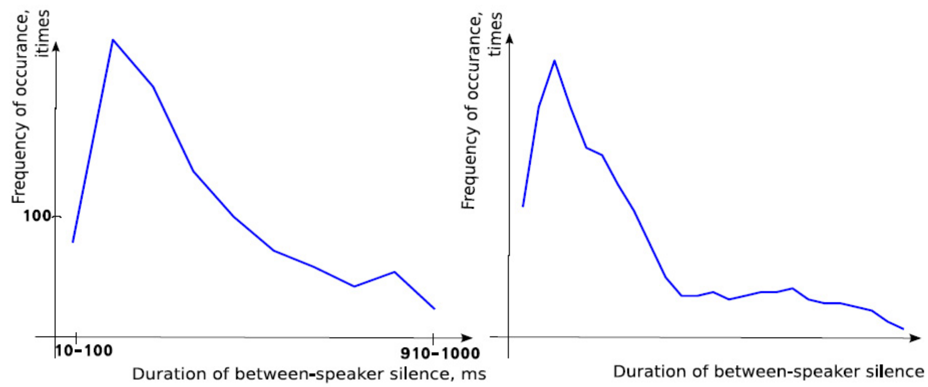

Fig. 4. Left: A histogram of the durations of silences between turns observed experimentally in actual human conversation is drawn from the data provided in (M. Wilson & Wilson, 2005). Right: A corresponding histogram generated from the dynamic turn taking model. As the time scale in our model is arbitrary, we compare only the overall form of the distributions of the silent periods' durations. See Fig. 6 for how silences are measured in the model.

4 A Dynamic Field Model of Turn Taking

Generating the amazingly smooth and fast transitions between turns requires that inherently time-variant events, the individual speaking turns, be coordinated with good temporal precision. Some authors have postulated that turns are governed by oscillators, so that their durations are multiples of a base unit [7]. This does not seem to capture the full level of temporal flexibility of embodied communication as reflected, in fact, by the distributions of silent intervals at turn switches observed experimentally by these very authors (Fig. 4, left). These distributions have a lot of counts at surprisingly short times (of the order of 100 milliseconds!) but are still broad relative to their means.

Another view is that turn taking requires anticipation, that is, predicting when a chance to switch may occur [8]. Here we present a dynamical field model that is primarily a metaphor for how the timing properties of turn taking may arise. Within this picture, there is no fixed underlying oscillator, although the duration of each act is timed and may thus be predictable. The model does not have an explicit mechanism for anticipation, but turn switching is based on receiving a graded signal from the communication partner. This may, in effect, generate anticipation. The model does not address the rich inner structure of dialog, nor the multi-modality of its embodiment, and thus remains at a much more metaphorical level than has typically been the case for DFT based models. Hopefully, the model can provide a perspective for how the framework of dynamical systems thinking may help understand the autonomous, graded, and real-time structure of embodied communication.

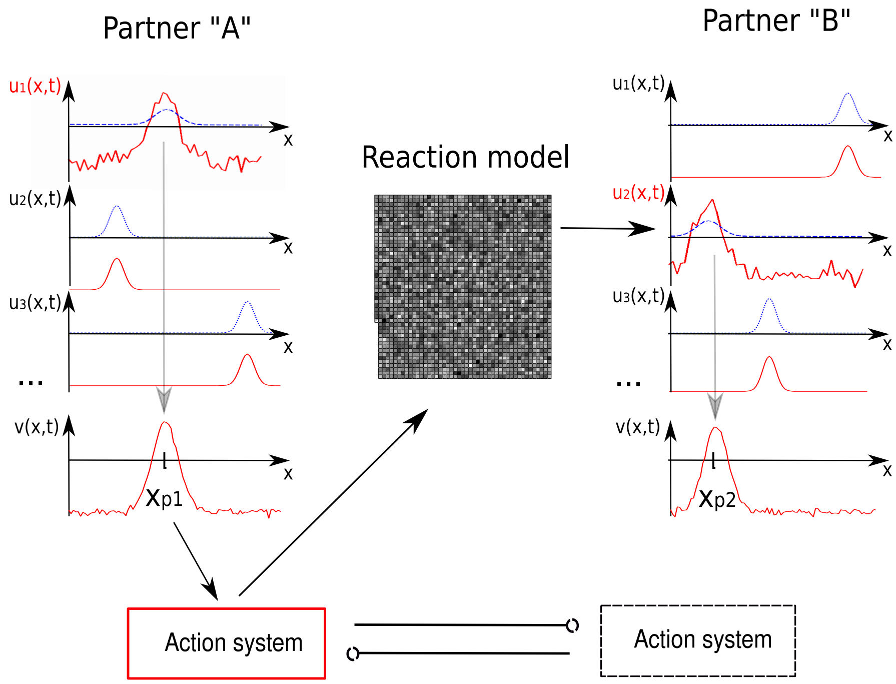

Fig. 5. The dynamic field model of turn taking consists of a dynamic sequence generation model for each of two communication partners "A" and "B" (stacks of fields on the left and the right). The action system itself, which generates communicative acts, is modelled by a neuronal oscillator, which generates a single timed event when activated and then deactivates itself. The action systems of the two partners compete so that only is active at a time. Another channel of interaction is the reaction model which provides localized input into the sequence generation system of one partner (at the depicted instant: "B") based on the current action of the other partner (at the depicted instant: "A").

We sketch the conceptual structure of the model. The mathematical equations are listed in the Appendix. Consider two partners, "A" and "B", communicating with each other (Fig. 5). Each actor is modelled as a sequence generation system defined over a metric dimension, x, that represents a feature value that characterizes each communicative act. For our simple toy model, this feature dimension was simply the planned duration of each communicative act.

There are two contributions to the contents of the sequence. One represents a prior plan of a series of communicative acts, modelled by localized inputs associated with each step in the sequence. The other is a reactive component, represented by localized input generated at each turn by the action of the other partner. Presently, this is merely a random mapping from the feature value of that other partner's action onto a feature value of the present partner's sequence plan.

Assume, for now, that communicator "A" has generated the beginning of a planned sequence, so that a peak in ordinal field number 1 is located over the associated input (left panel of Fig. 5). This leads to a matching peak at location, x_{p1}, in that actor's motor field (bottom of the stack). Communication partner "B" might be in another state. This communicator may have no peak at all (no communicative intention). More typically, however, this partner may have some communicative intention represented by a peak at a particular location in a particular ordinal position of a prior communicative plan. In the figure, partner "B" is at the second ordinal position in an ongoing sequence having generated a peak in the second ordinal field together with the matching peak at location, x_{p2}, in the motor field. Thus, communicator "B" faces the action of "A" already with some communicative intention, a prepared communicative act.

We model turn taking by providing a caricature of the action system that each communicator uses to "act out" communicative intentions. These systems generate single "actions" with a well-defined duration by starting a limit cycle oscillator with a specific frequency. In fact, the frequency of the oscillator and thus the duration of its "action" is encoded along the metric dimension of the sequence generation systems. Thus, different actions take different amounts of time. Each oscillator has two inner state variables (excitation and inhibition). Excitation arises from zero, reaches a maximum, and falls back to zero for each cycle of the oscillator. After a single cycle, the oscillator turns itself off. This is controlled by an activation variable, which is either in an "on" state or an 'off" state. Input from the motor field and absence of competition from the other actor turns the activation variable "on". When the associated oscillator reaches the end of a cycle, then the activation variable is switched to an "off" state through a dynamic instability.

The level of the activation variables controlling each oscillator is exchanged between the two communicators as a signal for how close they are to yielding the turn. A large level of activation of action system "A" inhibits action system "B" and vice versa.

Figure 6 illustrates simulations of this complete model by showing for both actors the time courses of a state variable characterizing the respective oscillator. Episodes of oscillator activation are individual communicative actions. These episodes vary in duration as dictated by the frequency of the limit cycle oscillators encoded by the location of peaks in the neural fields of the sequence generation systems. An action system, whose state variable is below a threshold, is in the off state waiting its turn. When the state variables of both action systems are below threshold, then both action systems are in the off state and a period of silence between turns is observed. The amount of silence at each turn switch varies because the state variables change at variable rates due to the variable durations of actions. Moreover, turn switches involve instabilities in the dynamics of the action systems. This gives noise a sizeable influence on the exact time at which a transition is realized. As a result, the histogram of the durations

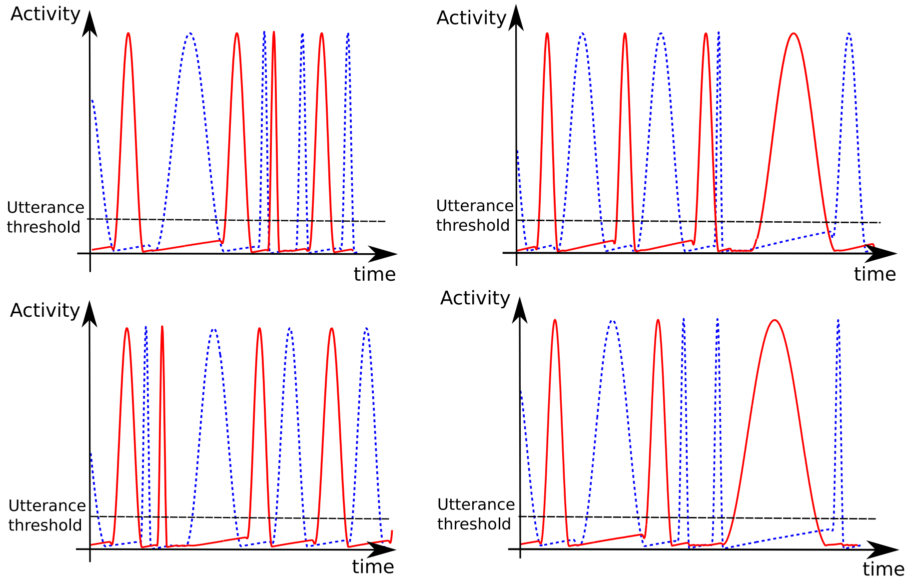

Fig. 6. Sample trajectories of turn taking generated by the model. The activity of each of the oscillators modelling the action systems of actor “A” (solid lines) and “B” (dotted lines) is shown by plotting one of two oscillator state functions. Positive values beyond a threshold (solid horizontal lines) signal that the oscillator is in the “on” state and a communicative act is ongoing. When the state variables of both actors are below the threshold, then this is interpreted as a silence interval the duration of which contributes to the histogram shown in Fig. 4.

of silent intervals obtained from an ensemble of simulation runs shown in Fig. 4 is quite broad, although centered on a most frequent interval. This characteristic of a small mode and long tail matches the shape of the distributions obtained from human data shown in the same figure on the left.

5 Discussion

We presented a simple mathematical model of turn taking as a metaphor for how the time structure of embodied communication could be understood within the neuronally based theoretical framework of Dynamic Field Theory. The dynamic activation fields and dynamic action systems in the model generate their time courses *autonomously*, based on continuous time. They are not paced by a rigid input-compute-output cycle, but are open to sensory input at any time. Because the systems are almost always in a *stable* state, such continuous coupling does not prevent the systems from performing their assigned function. The systems are sensitive to input only while near an instability, which leads to a switch of turn.

These instabilities start at the level of the action system, which turn on and off in response to graded signals received internally as well as from the communication partner. The instabilities amplify small graded changes of signals into macroscopic changes. This is how dynamical systems can make sense of and depend on *graded* variables while at the same time being able to make categorical decisions and to discard graded information as they set on the course of a new action step.

The simplistic models of the action systems as self-controlled, one-shot oscillators stands in for the much richer action systems engaged when people communicate. These include the speech articulatory system, systems controlling the prosody of their utterances, gesturing, controlling facial expression and body posture. All these systems may have graded components, from which communication patterns may derive signals for how close the actor is to yielding the turn. That we model such complex systems as stable limit cycle oscillators is a deliberate scientific move. The coupling among stable oscillators is the basis for coordination of timed actions [34, 35]. Thus, it is easy to imagine how the multiple action involved in generating communicative acts maybe coordinated with each other as well as across two communication patterns through couplings of the kinds modelled in simplified form here. This metaphor may provide an avenue toward an account for the remarkable temporal regularity and order observed in embodied communication [36].

The meaning transmitted in and expressed through communicative acts has been modelled only minimally here. In DFT, meaning is encoded through continuous metric dimensions. The location of peaks along such dimensions signifies specific instances of such metric information. In principle, coupled networks of dynamic fields may generate distributed representations of perceptual objects (see [37] for an example). The multiple modalities and perceptual dimensions that are relevant to embodied communication could easily be cast in dynamic field terms. Within the toy model presented here, the field dimension represented the duration of a planned communicative act and in that respect acted merely as a place holder for more substantive communicative meaning.

The field dynamics provides a framework for integrating different sources of specification of meaning. For instance, a memory trace of previously activated or stimulated patterns of activation may bias the ordinal fields to particular locations. On the other hand, current input from the other actor's communicative act may be overlaid and fused with such a prior plan. This may lead to the selection of activation patterns that in some way match the received message. The metrics of patterns of activation within continuous activation fields forms the basis for determining such matches. This metrics can be exploited by conceiving of structured forward mappings, like those of connectionist networks, to replace the random world matrix of our toy model.

Thus, much of the machinery needed to enrich the processing exists within connectionist and dynamical systems approaches. The challenge will be to translate insights obtained from substantive models of the information processing involved in embodied cognition [8] into dynamic terms. What is needed for this to happen is that aspects of embodied communication that provide entry points into dynamical systems thinking are identified and collaboratively explored in both theory and experiment. During the research year, from which this book emerged, first steps in such a direction were made. We are looking forward to the new and exciting ideas that may emerge from this ongoing research effort.

A Mathematical Description of the Models

The sequence generation model. The dynamic fields, $u_i^A(x,t)$, represent the ordinal position, $i = 1, \ldots, N$, of items in the sequence of communication partner, A, along the feature dimension, x, and evolve in time, t, according to:

$$\begin{aligned}\tau_o\ \dot{u}_i^A(x,t) &= -u_i^A(x,t) + h_o^A + \int f\big(u_i^A(x',t)\big) w_{oo}(x,x') dx' \\ &+ C_+ F_{\text{env}} \int f\big(v^A(x',t)\big) f\big(u_{i-1}^A(x',t)\big) dx' \\ &- C_- \int f\big(u_{i+1}^A(x',t)\big) dx' + P_i^A(x,t) \\ &+ \int \text{WorldModel}(x,x') f\big(v^B(x',t)\big) dx'. \end{aligned} \tag{1}$$

The analogous equation for communication partner, B, is obtained by switching upper indices A and B. The symbols represent the following:

- τ_o: the time constant of the field dynamics;
- h_o^A: constant resting level of the field;
- $f(u) = 1/(1+\exp(-\beta(u-u_o))$: a sigmoidal function, where β is a parameter and u_0 a threshold;
- $w_{oo}(x,x') = -w_{\text{inhib}} + w_{\text{excite}} \exp(-(x-x')^2/2\sigma^2)$: the "mexican-hat" interaction kernel with parameters w_{inhib}, and w_{excite};
- $C_+ F_{\text{env}}$: models sensori-motor feedback about accomplishment of the current action, defined below;
- C_-: strength of backward inhibition along the stack;
- $P_i(x,t)$: localized preactivation of the ordinal field number i encoding what is represented or planned at that ordinal position in the sequence;
- WorldModel(x,x'): a $N \times N$ matrix models communication by associating an output of partner B within an input to partner A's ordinal stack. Here a random matrix.

The motor field, $v^A(x,t)$, of partner A is governed by:

$$\begin{aligned}\tau_M \dot{v^A}(x,t) &= -v^A(x,t) + h_m^A + \int f\big(v^A(x',t)\big) w_{mm}(x,x') dx' \\ &+ \Sigma_{i=0}^N \Big[\int f\big(u_i^A(x',t)\big) w_{mo}(x,x') dx' + C_+ \int f\big(u_i^A(x',t)\big) dx' \Big] \end{aligned} \tag{2}$$

Again, the analogous equation applies to communication partner, B. The new symbols represent the following:

- h_m^A: resting level;
- $w_{mm}(x, x')$: interaction kernel analogous to $w_{oo}(x, x')$;
- $w_{mo}(x, x')$: a gaussian projection kernel from the ordinal position stack to the motor field.

The turn taking model. The sensori-motor system that generates a communicative act is described by two action variables, x_A, and y_A. In the present model the only significance of these action variable is to signal that an action is ongoing. This happens when the associated dynamics has a limit cycle solution. The Hopf normal form generates such a limit cycle:

$$\begin{aligned} \tau_h \dot{x_A} &= g_A \left[\gamma(\mu - x_A^2 - y_A^2)x_A - \omega_A y_A\right] - (1 - g_A)x_A \\ \tau_h \dot{y_A} &= g_A \left[\gamma(\mu - x_A^2 - y_A^2)y_A + \omega_A x_A\right] - (1 - g_A)(y_A - 1) \end{aligned} \tag{3}$$

This is the equation for partner A. The analogous equation applies to partner B. The symbols mean the following:

- τ_h: time scale of the oscillator dynamics;
- γ: parameter determining the relaxation rate of the limit cycle;
- μ: parameter determining the amplitude ($= 2\sqrt{\mu}$) of the limit cycle;
- g_A (and the analogous g_B): dynamic neuronal activation variables described below that turn on ($g_A = 1$) and off ($g_A = 0$) the stable limit cycle solution. When they are off ($1 - g_A = 1$), the dynamics has a fixed point at ($x_A = 0, y_A = 1$).
- ω_A determines the frequency of the limit cycle and is determined from the motor field by
$$\omega_A = C \int x f(v^A(x,t))dx \tag{4}$$
where the constant C is a normalization factor.

The neuronal activation variables g_A and g_B evolve according to a competitive dynamics that is built from two normal forms of the degenerate pitchfork bifurcation, coupled competitively:

$$\tau_g \dot{g}_A = \alpha_A g_A - |\alpha_A| g_A^3 - g_B^2 g_A + \xi \tag{5}$$
$$\tau_g \dot{g}_B = \alpha_B g_B - |\alpha_B| g_B^3 - g_A^2 g_B + \xi \tag{6}$$

The coupling makes that only one neuron may be activated at a time, e.g., $g_A = 1$ and $g_B = 0$ (see [38] for mathematical analysis). The factors α_A or α_B determine which of the two neurons is activated. These factors are designed to be positive (enabling the associated neuron to be turned on) only if there is a peak in the motor field of the corresponding communication partner. They become very small when the associated oscillator is near one end of its limit cycle (e.g., $x_A \approx -\sqrt{\mu}$), generating a tendency for the associated oscillator to be turned off. These factors α_A and α_B also serve to generate the signal, F_{env}, that provides the "condition of satisfaction" to the stack of ordinal fields. Because the competitive dynamics go through a degenerate pitchfork bifurcation each time an oscillator is turned on or off, noise ξ is essential to push the neurons away from unstable states.

References

1. Thelen, E.: Time-scale dynamics and the development of an embodied cognition. In Port, R.F., van Gelder, T., eds.: Mind as motion: Explorations in the dynamics of cognition, Cambride, MA, MIT Press (1996) 69–100
2. Clark, A.: An embodied cognitive science. Trends in Cognitive Sciences **3**(9) (1999) 345–351
3. Anderson, M.L.: Embodied cognition: A field guide. Artificial Intelligence **149** (2003) 91–130
4. Schöner, G.: Dynamical systems approaches to cognition. In Sun, R., ed.: Cambridge Handbook of Computational Cognitive Modeling, Cambridge, UK, Cambridge University Press (2007)
5. Spencer, J.P., Schöner, G.: Bridging the representational gap in the dynamical systems approach to development. Developmental Science **6** (2003) 392–412
6. Sacks, H., Schegloff, E.A., Jefferson, G.: A Simplest Systematics for the Organization of Turn-Taking for Conversation. Language **50**(4) (1974) 696–735
7. Wilson, M., Wilson, T.P.: An oscillator model of the timing of turn-taking. Psychonomic Bulletin and Review **12**(6) (2005) 957–968
8. Thórisson, K.R.: Natural turn-taking needs no manual: Computational theory and model, from perception to action. In Granström, B., House, D., Karlsson, I., eds.: Multimodality in Language and Speech Systems, Dordrecht, The Netherlands, Kluwer Academic Publishers (2002) 173–207
9. Thelen, E., Schöner, G., Scheier, C., Smith, L.: The dynamics of embodiment: A field theory of infant perseverative reaching. Brain and Behavioral Sciences **24** (2001) 1–33
10. Hock, H.S., Schöner, G., Giese, M.A.: The dynamical foundations of motion pattern formation: Stability, selective adaptation, and perceptual continuity. Perception & Psychophysics **65** (2003) 429–457
11. Rumelhart, D.E., Norman, D.A.: Simulating a skilled typist: A study of the skilled motor performance. Cognitive Science **6** (1982) 1–36
12. Houghton, G.: The problem of serial order: a neural network model of sequence learning and recall. In Dale, R., Mellish, C., Zock, M., eds.: Current research in natural language generation, London, Academic Press Professional, Inc. (1990) 287–319
13. Boardman, I., Bullock, D.: A neural network model of serial order recall from short-term memory. In: Proceedings of the 1991 International Joint Conference on Neural Networks, July 8-12, Seattle WA, International Neural Network Society (1991) II–879–884
14. Beiser, D.G., Houk, J.C.: Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. Journal of Neurophysiology **79**(6) (1998) 3168–3188
15. Farrell, S., Lewandowsky, S.: An endogenous distributed model of ordering in serial recall. Psychonomic Bulletin and Review **9**(1) (2002) 59–79

16. Deco, G., Rolls, E.T.: Sequential memory: A putative neural and synaptic dynamical mechanism. Journal of Cognitive Neuroscience **17**(2) (2005) 294–307
17. Erlhagen, W., Schöner, G.: Dynamic field theory of movement preparation. Psychological Review **109** (2002) 545–572
18. Grossberg, S.: Biological competition: Decision rules, pattern formation, and oscillations. Proceedings of the National Academy of Sciences (USA) **77** (1980) 2338–2342
19. Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik **13** (1973) 55–80
20. Amari, S.: Dynamics of pattern formation in lateral-inhibition type neural fields. Biological Cybernetics **27** (1977) 77–87
21. Wilson, H.R.: Spikes, Decisions, and Actions: Dynamical Foundations of Neurosciences. Oxford University Press (1999)
22. Deco, G., Schürmann, B.: Information Dynamics: Foundations and Applications. Springer Verlag, New York, NY (2000)
23. Goldstone, R.L.: Similarity, interactive activation, and mapping. Journal of Experimental Psychology: Learning, Memory, and Cognition **20** (1994) 3– 28
24. McClelland, J.L., Rogers, T.T.: The parallel distributed processing approach to semantic cognition. Nature Reviews Neuroscience **4**(4) (2003) 310–322
25. Churchland, P.S., Sejnowski, T.J.: The computational brain. Bradford Book/The MIT Press, Cambridge, MA (1992)
26. Williams, R.J.: The logic of activation functions. In Rumelhart, D.E., McClelland, J.L., the PDP research group, eds.: Parallel distributed processing. Volume 1. (1986) 423–443
27. Grossberg, S.: The quantized geometry of visual space: The coherent computation of depth, form, and lightness. Behavioral and Brain Sciences **6** (1983) 625–692
28. Wilimzig, C., Schöner, G.: How categorical behavior emerges from continuous neural representations: Dynamic field theory. (in preparation)
29. Humphreys, G.W., Forde, E.M.E., Francis, D.: The organization of sequential actions. In Monsell, S., Driver, J., eds.: Control of Cognitive Processes — Attention and Performance XVIII, Cambridge, MA, MIT Press (2000) 427–442
30. Sandamirskaya, Y., Schöner, G.: Dynamical field theory of sequence generation. (in preparation)
31. Searle, J.R.: Intentionality — An essay in the philosophy of mind. Cambridge University Press (1983)
32. Aldridge, J.W., Berridge, K.C.: Coding of serial order by neostriatal neurons: A "natural action" approach to movement sequence. Journal of Neuroscience **18**(7) (1998) 2777–2787
33. Procyk, E., Tanaka, Y.L., Joseph, J.P.: Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. Nature Neuroscience **3** (2000) 502–508
34. Schöner, G., Kelso, J.A.S.: Dynamic pattern generation in behavioral and neural systems. Science **239** (1988) 1513–1520
35. Schöner, G.: Timing, clocks, and dynamical systems. Brain and Cognition **48** (2002) 31–51
36. Streek, J.: Gesture as communication i: Its coordination with gaze and speech. Communication Monographs **60**(4) (1993) 275–299
37. Johnson, J.S., Spencer, J.P., Schöner, G.: A dynamic neural field theory of multi-item visual working memory and change detection. In: Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci 2006), Vancouver, Canada (2006) 399–404

38. Schöner, G., Dose, M.: A dynamical systems approach to task-level system integration used to plan and control autonomous vehicle motion. Robotics and Autonomous Systems **10** (1992) 253–267