NEUROBIOLOGICAL INTERPRETATION OF A QUANTITATIVE TARGET APPROXIMATION MODEL FOR SPEECH ACTIONS

Bernd J. Kröger[1], Peter Birkholz[1], Jim Kannampuzha[1], Cornelia Eckers[1], Emily Kaufmann[2], Christiane Neuschaefer-Rube[1]

[1]Department of Phoniatrics, Pedaudiology, and Communication Disorders, RWTH Aachen University, Aachen, Germany

[2]Human Technology Centre, RWTH Aachen University, Aachen, Germany

{bkroeger, pbirkholz, jkannampuzha, ceckers, cneuschaefer}@ukaachen.de, kaufmann.emily@gmail.com

Abstract: No quantitative specifications are known for neurobiologically motivated motor plans of speech actions (i.e. syllables, words, or utterances). This paper is motivated by the notion that quantitative parameters – as they can be estimated using a three-parameter model of movement trajectory approximation – are valuable candidates for specifying each movement action within the motor plan of an entire speech action. This paper will also argue that each movement action is comprised of a preparation phase, a target approximation phase, a target phase, and a release phase. **Index Terms**: speech action; motor plan; movement action; vocal tract action; movement trajectory; target approximation

1 Introduction

One of the main goals of our group is to develop a neurobiologically based, comprehensive, quantitative model of speech production, perception and acquisition [1], [2], [3], [4], and [5]. Within this approach, *speech movement actions* are assumed to be the basic units which comprise syllables, words, and utterances [2]. Speech movement actions (e.g. a bilabial closing action for producing a /b/ or a bilabial closing action together with a glottal opening action for producing a /p/, etc.) are temporally organized in a *motor plan* or *action score* ([6] and see below). In neurobiologically based models of speech production, motor plans – i.e. representations of motor and articulatory movement planning for the *whole set of vocal tract articulators* (tongue, lips, lower jaw, velum, glottis, lungs) for whole syllables, words, or utterances – are assumed to be located within premotor cortical areas and within the anterior insula, while neural activations in primary motor areas are responsible for action execution (e.g. [7] and [8], p. 17). Current neurobiologically based models which describe speech movement actions focus mainly on the execution of single movement actions or simple sequential orderings of two or three movement actions, usually acting on the same articulator; e.g. [9], [10], [11], and [12]. Thus, on the one hand, there is so far no neurobiologically based quantitative model which is capable of generating comprehensive speech motor plans, taking into account the entire complexity of temporally overlapping speech movement actions, i.e. temporally overlapping vocalic tract-forming, consonantal closing or constriction-forming, velopharyngeal opening, and glottal opening or closing actions, which occur within a syllable, word, or utterance. But on the other hand, *elementary quantitative dynamic models* which describe target-directed movement trajectories using a small set of parameters do exist (for speech movement actions see [13], [14], [15], [16], [17], [18], and [19]). The primary goals of this paper are to offer a neurobiologically motivated interpretation of the parameters defined within our quantitative model describing movement trajectories ([18] and [19]) and to specify a feasible set of neurobiologically motivated parameters for a quantitative specification for speech motor plans on this basis.

2 Entire actions and movement actions

Each *entire action* can be defined as a specific *goal-directed behavior* of a subject [2]. Two major types of actions are *private actions*, such as grasping an object or walking, and *communicative actions*, such as speaking, manual gesturing, or producing facial expressions [ibid.]. In the case of a private action, the goal of an entire action may be to move an external object or to move the subject's own body towards a desired target location. In the case of speech, the goal of an entire action (also called *speech action*) is to produce an understandable word or utterance. A private grasping action may be composed of a series of *(basic) movement actions* which are *target-directed*, e.g. (i) reaching out towards the object with the hand-arm systems, (ii) grasping the object with the hands, and (iii) moving that object towards a desired target location. These movement actions are temporally organized in an *action score*. A speech action (e.g. the production of a word) is composed of a score of *speech movement actions* (also called *vocal tract actions*, see [6]). Vocal tract actions can be separated into four types of actions: supraglottal vocalic, supraglottal consonantal, velopharyngeal, and glottal. The goal of a vocalic or consonantal action is to produce a specific supraglottal vocalic vocal tract shape, consonantal constriction, or consonantal closure. The goal of a velopharyngeal action is, for example, to produce a velopharyngeal opening as part of a nasal sound. The goal of a glottal action is, for example, to produce a glottal closure as part of a voiced sound, or to produce a glottal opening as part of a voiceless sound [6]. An example for the temporal organization of speech movement actions by an action score is given in Fig. 1 for the German word "Kamm" ('comb'). It should be noted that each movement action may control more than one articulator (e.g. lower jaw *and* lips in the case of a bilabial closing action or lower jaw *and* tongue in the case of vocalic movement actions; see Fig. 2).

Thus, an *entire action* – defined above as a high-level goal-directed behavior – is implemented as a *unit for controlling an ensemble of articulators (or effectors) accomplishing a set of target-directed movements called (basic) movement actions*. As already stated above, the temporal relations of all speech movement or vocal tract actions comprising a speech action (e.g. a word) are specified within the action score. A time interval can be defined for the execution of the entire speech action (e.g. the whole word) as well as for each movement action of which this entire action is composed. Each time interval of a movement action can be divided into a preparation, approximation, target, and release phase. If only the execution of the movement action is taken into account, the preparation phase can be left out. Strong articulator *movement* within a speech movement action occurs during its approximation and release phase (dark blue rectangles in Fig. 1). Less movement occurs during the target phase because the spatial target has nearly been reached at this point (e.g. the time interval of closure in a consonantal closing action represents the target phase of a speech movement action). No articulator movement is generated from a speech movement action during its preparation phase. Thus, the preparation phase occurs during the time interval starting with the activation of the first motoneurons (i.e. motoneuron recruitment) for that speech movement action and ending with the onset of the articulator movement which is induced by this speech movement action (see also the next paragraph).

A target phase is allowed to be absent for a movement action in stress-timed languages, for example in the case of the realization of an unstressed (reduced) vowel. Here, vowel quality perception mainly results from perceptual processing of the formant transitions, which occur during the approximation phase of the movement action. A target phase may also be absent in the case of approximants, i.e. speech sounds which are defined mainly by articulator *movements* and which lack sustained constrictions or closures as seen in plosives or fricatives. But even in the case of plosives and fricatives, target phases may be strongly reduced if the speaking rate is fast and/or the speaking style is casual. Moreover, in the case of plosives and

fricatives the movement phase of the speech movement action is very important (i) for coding the place of articulation on the basis of formant transitions, and (ii) for indicating the closure or constriction.

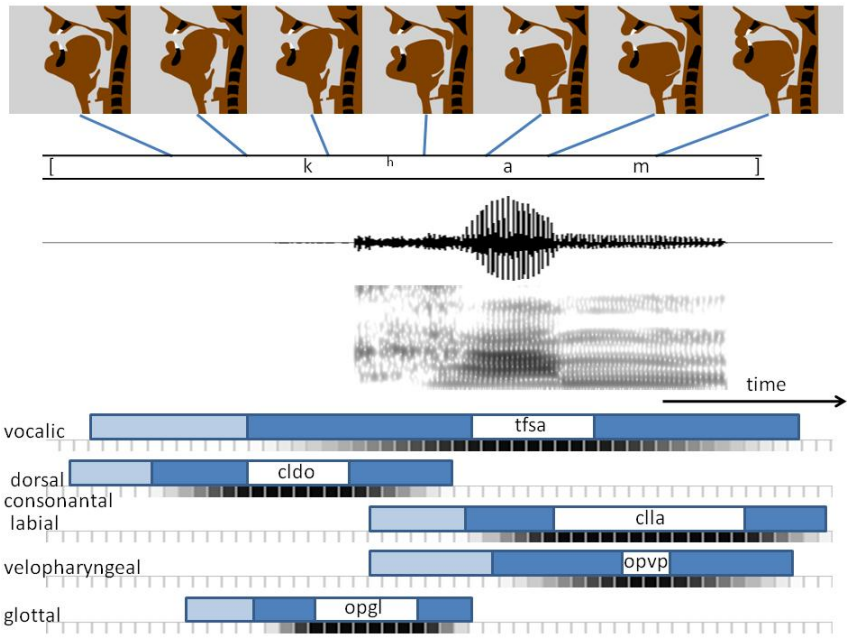

Figure 1 – Action score and part of the motor plan representation for the German word "Kamm" ('comb'). From top: midsagittal views of the vocal tract for seven points in time; phonetic transcription; oscillogram and spectrogram of the acoustic signal; and action score of the word /kam/. The action score is organized into four tiers: vocalic, consonantal, velopharyngeal, and glottal. The upper part of each tier indicates four time intervals for each vocal tract action: the preparation (light blue rectangle), approximation (dark blue rectangle), target (white rectangle), and release (dark blue rectangle) phases. Primary articulator movements occur within the approximation and release phases. The name of each vocal tract action is abbreviated using four letters which are displayed in the white rectangle (target phase) of each vocal tract action. Abbreviations are explained in Fig. 2 (below). The lower part of each action score tier indicates rows of motor plan neurons; white = no activation, black = full activation of a neuron. Each motor plan neuron represents a time interval of 12.5 ms, and each row of motor plan neurons indicates the degree of target approximation as it is reached by a vocal tract action at a specific point in time. (It should be noted that only the approximation and target phases of speech movement actions are displayed in earlier publications; e.g. [2]; also, the "approximation phase" was labeled "movement phase" in earlier publications, including [20].)

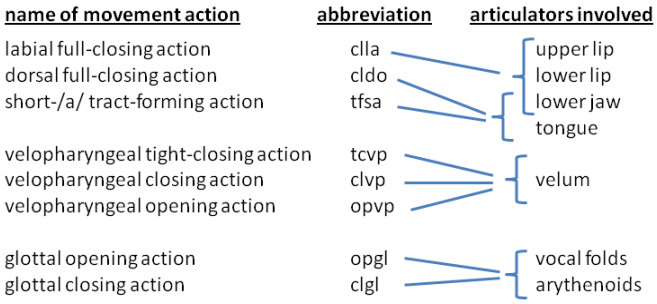

Figure 2 – List of vocal tract actions comprising the word /kam/ (Fig. 1) and list of articulators involved in the execution of the different vocal tract actions.

3 The neurobiological perspective on speech actions

From the neurobiological point of view we can differentiate four levels of action realization.

(i) At the *cognitive level*, an action activates its *cognitive symbolic representation.* This representation is a qualitative specification of the goal-directed behavior – e.g. to produce an understandable word – and thus it is closely related (a) to the meaning of a word or utterance (i.e. its semantic specification) as well as (b) to a qualitative representation of the goal-directed behavior of that action (e.g. a phonological specification of all movement actions comprising the speech action; for the distinction between cognitive and sensorimotor aspects of actions, see [2]. These symbolic action and movement action representations occur at the level of the mental lexicon, located within the temporal cortical lobe [21].

(ii) At a *high sensorimotor level*, the preparatory network for speech actions has been substantiated (preparatory loop, see [7]). This network – comprising the medial and dorsolateral premotor cortex, the anterior insula, and the superior cerebellum – is responsible for *motor planning*. A *motor plan* can be interpreted as a neurobiological realization of an action score. It comprises a coarse quantitative specification of the temporal relations between all movement actions, i.e. how the movement actions are coordinated with each other in the temporal domain (Fig. 1), but it does not determine the articulatory trajectories of movement actions in detail (see [8]).

The temporal delay between preparing an action (i.e. peak activation at the level of the prefrontal cortex for action planning) and activating motoneurons for its (first) movement actions (i.e. activating neurons at the level of the primary motor cortex for action execution) is about 100 to 200 msec. This value was estimated from manual actions (see [22]), but it should be noted that manual actions as well as speech actions are planned in the same premotor area (i.e. Brodmann's area 44) and that both types of actions are acquired by associative learning during "babbling" and "imitation" phases in early infancy (see [2]).

(iii) At a *lower sensorimotor level*, the executive network for speech actions has been substantiated (executive loop, see [7]). This network – comprising parts of the sensorimotor cortex including the primary motor cortex (with its upper motoneurons), basal ganglia, and inferior cerebellum – is responsible for the *execution* of the motor plan. At its lowest level, this executive network is linked with the lower motoneurons, which are located within the brain stem in the case of the vocal tract articulator system. Muscle forces are generated from motoneuron activations and result in articulator movements. Motoneurons can be grouped into specific sets which control specific muscles, parts of muscles or groups of muscles. Such sets of motoneurons and their corresponding muscles, parts or groups of muscles have been identified for different speech movement actions in [9], [10], [11], [12], and [23]. Moreover, these authors have adapted the *equilibrium point hypothesis* (i.e. *λ-model* of motor control) in order to model motoneuron activation for speech actions in a quantitative way. It this approach, speech movements are seen to be controlled by an ongoing activation of the motoneurons of several muscles or parts of muscles over the whole duration of a movement action. Co-contraction of agonist/antagonist muscle pairs is important in order to produce stable articulator movements; see especially Dang and Honda [11] for an excellent description of tongue dorsum, tongue tip, and lower jaw movements. Motoneuron activation results (a) from efferent higher level neural signals (control commands from premotor areas) and/or (b) from afferent proprioceptive neural feedback signals ([9] and [24]). The latter mainly provide information concerning the current muscle length and its time derivative [9].

It should be noted that movement actions may involve more than one articulator, and that most articulators are controlled by more than one muscle for executing a specific action; examples are given in Fig. 3. Thus, control of the execution of each movement action is a complex task comprising (i) a synergetic coordination of more than one articulator (see [14]) and (ii) a complex activation pattern for different muscles (or parts or groups of muscles),

which means it is also a complex activation pattern for different sets of motoneurons. The temporal coordination of the whole set of movement actions at the speech action level results in a complex higher-level premotor neural activation pattern during action execution. However, this does not represent the maximum level of complexity which may occur in speech production, since one articulator may be under the active control of two or more temporally co-occurring movement actions at a given time as well. An example is the control of the lower jaw in the word "Kamm" (Fig. 1), which is governed by the short-/a/ tract-forming action but also by (i) the dorsal closing action for /k/ at the beginning of the vocalic action (i.e. during its movement phase) and (ii) the labial closing action for /m/ at the end of the vocalic action (i.e. during its target phase).

The VITE model ([26] and [27]) not only takes into account the neural circuits comprising the lower motoneurons and the muscle fibers (as is the case in λ-model), but also includes the upper motoneurons, located at the level of the primary motor cortex, and it also includes parts of the cerebellum, basal ganglia, and thalamus as well as primary proprioceptive cortical areas, i.e. the pyramidal and extrapyramidal motor pathways. This model postulates ongoing activation of the primary motor areas during the entire execution of movement actions, as is proposed by the λ-model for the lower motoneuron level.

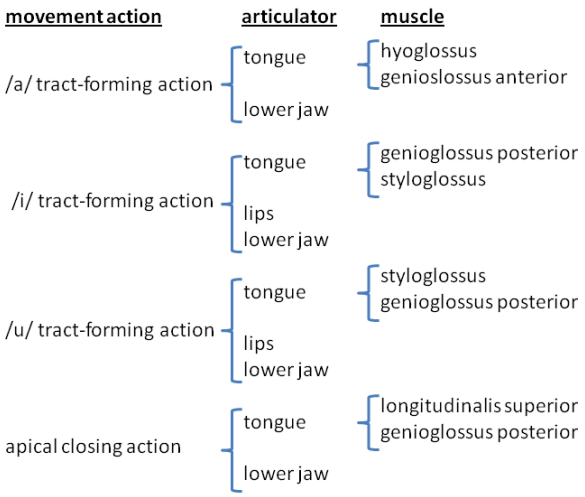

Figure 3 – List of three vocalic and one consonantal movement actions, controlling different articulators. Each articulator is controlled by a specific set of muscles for each movement action; muscles are specified here only for the tongue following [10], [11], and [25].

It should be noted that in addition to distinguishing between planning and execution, as introduced above, a network for "being prepared" and/or for "initiating an action" (i.e. a network which is functionally activated after planning/preparation and before execution) has recently been substantiated at the level of the supplementary motor area (see [28]).

(iv) At the *articulatory level*, the muscle force, muscle mass, and mass of bony structures as well as the damping characteristics of the whole *vocal tract articulator system* define the *dynamical behavior of each articulator* for each movement action. The resulting kinematics of all articulators during the execution of an action is the basis for action perception (see e.g. [2]). Because the vocal tract articulator system comprises mostly muscle and only few bony structures, mass is neglected in many approaches to speech articulation. Dang and Honda [11] completely neglect articulator mass in their dynamical approach and model the motoneuron–muscle system (i.e. the articulator system) using spring and damping elements. Thus, in this approach, not the muscle–articulator system, but rather the motoneuron–muscle system mainly determines the time delay for articulator movement after motoneuron activation. The

build-up of the muscular force after a step pulse activation of the motoneuron occurs gradually with a time delay of around 90 msec ([24], p. 376 and [9], p. 1618). A second order low pass filter may be used for modeling this gradual build-up of muscle force (ibid.).

Thus, the time delay from onset of action planning (i.e. onset of activation of premotor cortical areas e.g. for the preparation of a syllable or word) to onset of motoneuron activation for the first movement actions (*planning time for the entire action*) is about 100–200 msec (e.g. for the planning of a whole syllable; see above), and in addition the time delay between onset of neural activation for a specific movement action (i.e. onset of activation of primary motor areas) to onset of the intended articulator movement for a movement action (*movement execution delay*) is about 90 msec. There may also be an additional planning or preparation time interval for each movement action since different articulators must be coordinated synergetically for the execution of each movement action. This *preparation time for movement actions* occurs as part of the planning time for the entire speech action (e.g. a syllable) and must be added to the movement execution delay of 90 msec in order to specify the duration of the *preparation phase* of a movement action (light blue rectangles in Fig. 1). The preparation phase of a movement action is the time interval between the starting time for neural activation of a movement action (not the entire action!) and the starting time of articulator movements generated by that movement action.

4 Quantitative approaches to modeling speech movement actions

A preliminary quantitative concept which can describe speech movement actions is the *critically damped second-order dynamical system* (spring–mass system as a representation of an articulator–muscle system, e.g. [14], and [17]). In this approach, a time-invariant spatial *target* is assumed to represent the spatial target for each type of movement action. Also, a *stiffness* value determines the rapidity with which the target is approximated (the higher the stiffness level, the shorter the movement phase of that movement action); in the approach of Kröger et al. [17], the term 'stiffness' is replaced by the *strength of activation* of the force of a movement action. The force is assumed to be generated by the action, which induces muscle forces and initiates an attractor force, leading to target-directed articulator movements. It is an important feature of these models that the basic dynamic system is *time variant*. Each movement action is limited in time, i.e. there exists a point in time at which the movement action starts (is "switched on") and ends (is "switched off"). The details as to how the timing relations between movement actions should be quantified are still being debated. Saltzman and Byrd [15] describe a mechanism for the relative timing of movement actions based on the coupled oscillator assumption, i.e. based on *intrinsic* bodily processes. Kröger et al. [17] assume that timing relations for movement actions in speech are learned during the babbling and imitation phases of speech acquisition – in which imitation training especially can be seen as an *extrinsically* motivated behavior. This learning is governed by trial-and-error procedures since the improper timing of movement actions directly leads to mispronunciations and thus to a failure concerning the goal of speech actions, which is to produce understandable words and utterances. No attempts have been made by Kröger et al. [17] until now to identify *mechanisms* or *rules* for the timing or phasing of speech movement actions. But since the action scores of frequent syllables are assumed to be ordered in self-organizing maps in our approach (ibid.), generalization occurs and thus rule-extraction may occur as a byproduct (see e.g. [29] and [30]).

If the goal of a quantitative model for describing movement actions is to match natural movement trajectories with maximal accuracy, the assumption of the second-order dynamical system is only tenable if the activation of each action – i.e. the "switching on" and "switching off" of the attractor's force, defined by the second-order dynamical system – is modeled with a gradual or smooth onset and offset. This results, for example, in a six-parameter model for describing speech movement actions [17]. Four parameters describe starting and ending times

of onset and offset intervals for the force of the movement action; one parameter describes the maximal strength of the force, which occurs within the target phase of the movement action (Fig. 1); and one parameter describes the spatial target location. This six-parameter model is capable of describing the natural movement trajectories of speech articulators with high accuracy. Moreover, the onset and offset intervals, as well as the strength of the force, result from the parameter approximation procedure, while the spatial target locations of actions – defining the spatial target of the appropriate speech sounds – are predetermined.

In the same vein, a three-parameter model for describing articulatory movement trajectories has been developed by Birkholz et al. ([18] and [19]). This model is also time-variant and based on a dynamical systems approach. Up to this point, this approaches cited above have made the assumption of a second-order system based on the neurobiological and physical viewpoint. However, for the preparation/planning and execution of articulator movements during speech, this assumption is not justified. For this reason, Birkholz et al. ([18] and [19]) assume a higher order, i.e. a sixth to tenth order dynamical system which is "switched on" by a step pulse command. The model comprises only three parameters: a *command onset time*, a time constant for the *rapidity* of target approximation, and the *spatial target* location (ibid.). As in other approaches, the target location can be predicted, while command onset time and movement rapidity are estimated based on movement trajectory approximation for each movement action occurring within the entire speech action. Since the high order (sixth to tenth order) dynamical system leads to a time delay of approximately 50–100 msec or more between command onset and the onset of the resulting articulator movement (ibid., Fig. 4 and Fig. 10), this onset time of the dynamical system can be interpreted as the onset time for the neural activation of a movement action (preparation time of a movement action; see above). An abrupt step pulse for which marks the "switch on" of the dynamical system is introduced in this approach (ibid.) and does not lead to any disadvantage in terms of the how well it describes natural movement trajectories. Thus, in this approach the dynamical system may be interpreted as a comprehensive model for the preparation and execution of a movement action. But it should be noted that in its current version, the three-parameter model only aims for modeling CV or VV-sequences (C = consonant, V = vowel). No attempts are made to model more complex entire speech actions. Thus, in the current version of the three-parameter model, the dynamical system parameters change in a step-wise manner from the command onset time of a preceding movement action to the command onset time of the following movement action within a sequence of movement actions which act on the same articulator.

5 How can motor plans for speech actions be represented in a neurobiological model of speech production?

An important feature of our neurobiological model of speech production, perception, and acquisition is that it takes into account neural representations of motor plans as well as auditory and somatosensory representations of entire speech actions [1]. A preliminary neural representation of the temporal pattern of movement actions within a motor plan of speech is described in [4] and [31]. Here, each movement action is divided into an approximation, target, and release phase (cf. Fig. 1); a preparation phase was not included in earlier studies. An estimation of the temporal duration of these different phases of a movement action was done on the basis of acoustic and articulatory data (see [32] and [33]).

The *approximation phase* (called the onset phase in [32] and [33]) represents the first main movement phase of the action, i.e. the time interval between the starting time of the intended target-directed articulator movement and the point at which the target region is approximated. The duration of the approximation phase was found to be long for vocalic and velopharyngeal movement actions, and significantly shorter for consonantal and glottal movement actions (ibid.).

The *target phase* represents the time interval in which the spatial target is approximated. In the case of consonantal actions, the target phase represents the time interval of closure or constriction; in the case of vocalic actions, the target phase represents the short time interval or point in time in which the target vocalic vocal tract shape is maximally approximated. In the case of a velopharyngeal opening action, the target phase represents the relatively short time interval of maximal velopharyngeal opening. In the case of a glottal opening action, the target phase represents the relatively short time interval of maximal glottal opening; in the case of a glottal closing action, the target phase represents the time interval of glottal closure and thus of phonation.

The *release phase* (called the offset phase in [ibid.]) represents the second main movement phase of the action, i.e. the time interval between the release of the target and the end of the movement action. If the next movement action within a motor plan of a speech action directly involves that same articulator (i.e. is on that same articulatory tier; see Fig. 1), the release phase of the preceding movement actions is assumed to overlap completely in time with the approximation phase of the following movement action. If no further movement action directly follows on the same articulatory tier, it is assumed that the articulator moves back to a neutral position during the release phase, and that the release phase ends at that point in time at which the relevant articulator movement ends.

In earlier studies (e.g. [2]), the action concept comprises the approximation and target phase. In this paper, an additional release phase is introduced for each movement action. This release phase is introduced in order to account for the overlapping of movement actions seen in speech production. For example, in the case of consonantal movement actions, the release from consonantal constriction or closure need no longer be described as being controlled exclusively by the following (vocalic) movement action; another example would be the case of vocalic movement actions which influence the following (vocalic and/or consonantal) movement actions (carry-over or left-to-right coarticulation, e.g. [34]). In the same way, the approximation phase of a movement action can be interpreted as the time interval for allowing anticipatory or right-to-left coarticulation, e.g. [35]. It should be noted that in our action-based approach, both carry-over and anticipatory coarticulation result directly from movement action sequencing (cf. [36]) as well as from the temporal overlap of movement actions (cf. [37]).

Within the neural representation of the motor plan of an entire speech action ([4] and [31]), the neural activation occurring at each point in time during the execution of a movement action represents the instantaneous degree of target approximation realized by the movement action. Thus, neural activation increases during the approximation phase, reaches a peak during the target phase, and decreases during the release phase (see Fig. 1). Also, the neural representation should reflect the cortical premotor and primary motor activation which occurs during the preparation and the execution of a movement action. Thus, in addition to the information about target approximation, the temporal location of the *preparation phase* should also appear within the neural representation of the motor plan of an entire speech action. It can be seen in Fig. 1 that the duration of the preparation phase of a movement action depends on two factors: (i) the rapidity of the movement action (i.e. on the duration of the approximation phase) and (ii) the complexity of a movement action (e.g. a vocalic movement action requires more than one articulator, while a glottal movement action does not).

Furthermore, the inclusion of the preparation phase (as it is quantitatively described in the three-parameter model [18]) allows for more precise descriptions of the *on-line timing of movement actions*. For example, onset of labial closing action and the velopharyngeal opening action for the production of the nasal consonant in /kam/ nearly coincide in time (see Fig. 1). Also, the vocalic action for the short /a/ is already prepared during the target approximation phase of the dorsal closing action for /k/ in such a way that the vocalic movement action is

able to start at that point in time in which the consonantal closure is reached (Fig. 1). Thus, the break-down of each speech movement action into a preparation, approximation, target, and release phase (see Fig. 1) seems to be a feasible basis for coding the timing details of the whole ensemble of movement actions comprising a speech action.

In addition to this temporal information, other information is stored within the neural representation of each movement action: A motor plan describing an entire speech action represents all movement action parameters occurring in the quantitative three-parameter model [18]. This additional information concerns (i) the *spatial target* of the movement action, (ii) the *set of articulators involved* in the execution of a movement action and (iii) the *degree to which an articulator is involved* (see the dominance concept for articulatory movements as is mentioned in [6]). This additional information is not displayed in Fig. 1.

6 Discussion

It has been argued in this paper that parameters defined in a quantitative model – which describes articulatory movement trajectories (i.e. the three-parameter model [18]) – can be taken as a basis for generating speech motor plans. The movement action onset time quantified in that approach can be interpreted as the beginning of the preparation phase of a movement action. But it should be kept in mind that – in contrast to that quantitative approach [ibid.] – the assumption that the movement trajectory during the preparation phase of a movement action can be fully specified is not tenable. Planning or preparation as it occurs at cortical levels ends with a "crude representation of the intended motion" rather than with a "pre-computed command signal" leading to a precise movement trajectory which simply must be executed [8]. Rather, control of the movement action is ongoing throughout the whole time interval of action execution, i.e. during the approximation, target and release phases. Thus, neural activation for a movement action holds over the whole time interval of action execution as is exemplified in Fig. 1; i.e. the degree of target approximation during the execution of each movement action has to be monitored. Even for well-practiced movements, the role of on-line processes involving both afferent sensory feedback signals and efferent signals stemming from knowledge concerning learned internal forward models must be emphasized (see [8], p. 19).

In contrast to earlier second order dynamical models describing articulatory movement trajectories (e.g. [17]), the assumption of an abrupt onset of a higher level command for the preparation of a movement action is tenable in this approach. Neural signals (firing rates) often indicate abrupt changes. The step pulse command as introduced in the three-parameter model is accounted for by using the high order (i.e. sixth to tenth order) dynamical system for trajectory generation. Also, this high order of the dynamical system is responsible for the long time interval between the step pulse command and the onset of articulator movement.

Last but not least it should be kept in mind that quantitative models for describing movement trajectories (e.g. Birkholz et al. [18]) have no intrinsic neurobiological motivation. A goal of upcoming research projects should be to establish a neurobiologically-based quantitative approach for describing motor plans of entire speech actions as well as for describing the basic building blocks of speech motor plans, i.e. speech movement actions.

Acknowledgements This work was supported in part by the German Research Council, project KR 1439/15-1, and in part by EU-COST action 2102.

Literature

[1] Kröger, B.J., Kannampuzha, J., Neuschaefer-Rube, C., "Towards a neurocomputational model of speech production and perception", Speech Communication 51: 793-809, 2009.

[2] Kröger, B.J., Kopp, S., Lowit, A., "A model for production, perception, and acquisition of actions in face-to-face communication", Cognitive Processing 11: 187-205, 2010.

[3] Kröger, B.J., Birkholz, P., Kannampuzha, J., Neuschaefer-Rube, C., "Categorical perception of consonants and vowels: Evidence from a neurophonetic model of speech production and perception". In: A. Esposito, A.M. Esposito, R. Martone, V.C. Müller, G. Scarpetta [Eds], Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues. LNCS 6456 (Springer, Berlin), 354-361, 2011.

[4] Kröger, B.J., Birkholz, P., Kannampuzha, J., Kaufmann, E., Neuschaefer-Rube, C., "Towards the Acquisition of a Sensorimotor Vocal Tract Action Repository within a Neural Model of Speech Processing". In: A. Esposito et al. [Eds], Lecture Notes in Computer Sciences, in press.

[5] Kröger; B.J., Birkholz, P., Neuschaefer-Rube, C., "Towards an articulation-based developmental robotics approach for word processing in face-to-face communication", PALADYN Journal of Behavioral Robotics, in press.

[6] Kröger, B.J., Birkholz, P., "A gesture-based concept for speech movement control in articulatory speech synthesis". In: A. Esposito, M. Faundez-Zanuy, E. Keller, M. Marinaro [Eds], Verbal and Nonverbal Communication Behaviours, LNAI 4775 (Springer, Berlin), 174-189, 2007.

[7] Riecker, A., Mathiak, K., Wildgruper, D., Erb, M., Hertrich, I., Grodd, W., Ackermann, H., "fMRI reveals two distinct cerebral networks subserving speech motor control", Neurology 64: 700-706, 2005.

[8] Cisek, P., "Neural representations of motor plans, desired trajectories and controlled objects", Cognitive Processing 6: 15-24, 2005.

[9] Sanguineti, V., Laboisiere, R., Ostry, D.J., "A dynamic biomechanical model for neural control of speech production", Journal of the Acoustical Society of America 103: 1615-1627, 1998.

[10] Perrier, P., Payan, Y., Zandipour, M., Perkell, J., "Influences of tongue biomechanics on speech movements during the production of velar stop consonants: a modeling study", Journal of the Acoustical Society of America 114: 1582-1599, 2003.

[11] Dang, J., Honda, K., "Construction and control of a physiological articulatory model", Journal of the Acoustical Society of America 115: 853-870, 2004.

[12] Buchaillard, S., Perrier, P., "A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning", Journal of the Acoustical Society of America 126: 2033-2051, 2009.

[13] Saltzman, E., Kelso J. A. S., "Skilled actions: A task dynamics approach", Psychological Review 94: 84-106, 1987.

[14] Saltzman, E., Munhall, K. G., "A dynamical approach to gestural patterning in speech production", Ecological Psychology 1: 333-382, 1989.

[15] Saltzman, E., Byrd, D., "Task-dynamics of gestural timing: Phase windows and multifrequency rhythms", Human Movement Science 19: 999-526, 2000.

[16] Ogata, K., Sonoda, Y., "Reproduction of articulatory behavior based on the parameterization of articulatory movements", Acoustical Science and Technology 24: 403-405, 2003.

[17] Kröger, B.J., Schröder, G., Opgen-Rhein, C., "A gesture-based dynamic model describing articulatory movement data", Journal of the Acoustical Society of America 98: 1878-1889, 1995.

[18] Birkholz, P., Kröger, B. J., Neuschaefer-Rube, C., "Model-based reproduction of articulatory trajectories for consonant-vowel sequences", IEEE Transactions on Audio, Speech, and Language Processing, 19: 1422-1433, 2011.
[19] Birkholz, P., Kröger, B. J., Neuschaefer-Rube, C., "Articulatory synthesis and perception of plosive-vowel syllables with virtual consonant targets", Proceedings of Interspeech 2010 (Makuhari, Japan), 1017-1020, 2010.
[20] Kröger, B.J., Birkholz, P., Kaufmann, E., Neuschaefer-Rube, C., "Beyond vocal tract actions: Speech prosody and co-verbal gesturing in face-to-face communication", this volume.
[21] Hickok, G., Poeppel, D., "The cortical organization of speech processing", Nature Reviews Neuroscience 8: 393-402, 2007.
[22] Nishitani, N., Hari, R., "Temporal dynamics of cortical representation for action", Proceedings of the National Academy of Sciences of the United States of America, PNAS 97: 913-918, 2000.
[23] Perrier, P., Loevenbruck, H., Payan, Y., "Control of tongue movements in speech: the equilibrium point hypothesis perspective", Journal of Phonetics 24: 53-75, 1996.
[24] Laboissiere, R., Ostry, D.J., Feldman A.G., "The control of multi-muscle systems: human jaw and hyoid movements", Biological Cybernetics 74: 373-384, 1996.
[25] Honda, K., "Organization of tongue articulation for vowels", Journal of Phonetics 24: 39-52, 1996.
[26] Bullock, D., Cisek, P., Grossberg, S., "Cortical networks for control of voluntary arm movements under variable force conditions", Cerebral Cortex 8: 48-62, 1998.
[27] Bullock D., Grossberg S., "VITE and FLETE: Neural modules for trajectory formation and postural control". In: W.A. Hershberger [Ed.] Volitional Action (Elsevier, North-Holland), 253-297, 1989.
[28] Brendel, B., Hertrich, I., Erb, M., Lindner, A., Riecker, A., Grodd, W., Ackermann, H., "The contribution of mesiofrontal cortex to the preparation and execution of repetitive syllable productions: an fMRI study", NeuroImage 50: 1219-1230, 2010.
[29] Kohonen, T., Oja, E., Simula, O., Visa, A., Kangas, J., "Engineering applications of the self-organizing map", Proceedings of the IEEE 84: 1358-1384, 1996.
[30] Ritter, H., "Self-organizing semantic maps", Biological Cybernetics 61: 241-254, 1989.
[31] Kannampuzha, J. Eckers, C., Kröger B.J. "Training einer sich selbst organisierenden phonetischen Karte im neurobiologischen Sprachverarbeitungsmodell MSYL", this volume.
[32] Bauer, D., Kannampuzha, J., Kröger, B.J., "Articulatory Speech Re-Synthesis: Profiting from natural acoustic speech data". In: A. Esposito, R. Vich [Eds.] Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions, LNAI 5641 (Springer, Berlin), 344-355, 2009.
[33] Bauer, D., Kannampuzha, J., Hoole, P., Kröger, B.J. "Gesture duration and articulator velocity in plosive-vowel-transitions". In: A. Esposito, N. Campbell, N. Vogel, A. Hussain, A. Nijholt [Eds.] Development of Multimodal Interfaces: Active Listening and Synchrony, LNCS 5967 (Springer, Berlin), 346-353, 2010.
[34] Magen, H.S., "The extent of vowel-to-vowel coarticulation in English", Journal of Phonetics 25: 187-205, 1997.
[35] Bell-Berti, F., Harris, K.S., "Anticipatory coarticulation: Some implications from a study of lip rounding", Journal of the Acoustical Society of America 65: 1268-1270, 1979.
[36] Ostry, D.J., Gribble, P.L., Gracco, V.L., "Coarticulation of jaw movements in speech production: Is context sensitivity in speech kinematics centrally planned?" The Journal of Neuroscience 16: 1570-1579, 1996.
[37] Whalen, D.H., "Coarticulation is largely planned", Journal of Phonetics 18: 3-35, 1990.