PRODUCING SYLLABLES: MOTOR PLANNING, MOTOR PROGRAMMING AND EXECUTION

Bernd J. Kröger[1] & Trevor Bekolay[2]

[1]*Departiment of Phoniatrics, Pedaudiology, and Communication Disorders, Medical Faculty, RWTH Aachen University, Aachen, Germany*
bernd.kroeger@rwth-aachen.de

[2]*Applied Brain Research, Waterloo, ON, Canada*

Abstract: Speech production is a hierarchically organized process involving cortical, sub-cortical, and peripheral components. While the cognitive-linguistic processing levels are already investigated extensively this is not the case for the sensorimotor levels. We here introduce a neurobiologically inspired quantitative and computer-implemented production model in which four main modules can be separated: (i) linguistic planning (from concepts via lemmas towards phonological forms including syllabification), (ii) motor planning (establishing a raw gesture score), (iii) motor programming (establishing a fully specified gesture score including specification of muscle activation patterns for execution), and (iv) execution (i.e., generation of articulatory movements and of the acoustic speech signal). This quantitative computer-implemented production model is capable of producing all syllables of Standard German. The model allows a detailed description of motor planning, motor programming, and execution and it includes a detailed and quantitative concept for generating syllables by using speech movement units (also called gestures) as basic production units. This paper concentrates on detailed structuring of gestures, i.e., by calculating the contribution of different articulators in the realization of target-directed speech movement units realizing linguistically relevant vocal tract constrictions.

1 Background

Speech production as hierarchically organized process involves cortical, subcortical, and peripheral components and can be separated in cognitive-linguistic and sensorimotor processing [1, 2]. A widely accepted approach describing the cognitive-linguistic processing is described by [1]. One of the most detailed neurobiologically based sensorimotor approach for speech production is the DIVA model (Directions Into Velocities of Articulators, see [3-5]). Input for the sensorimotor component is a sequence of phonologically specified syllables which are converted into motor plans and/or programs. A still open question is whether *motor planning* and *motor programming* should be separated and if yes, how to define motor panning and motor programming (e.g., [2]). We argue for a clear separation of motor planning and motor programming and we give a quantitative formulation of planning and programming by postulating *vocal tract gestures* (also called *vocal tract movement units* or *speech movement units* see [6-8], but abbreviated here in the following simply as *gestures*) as the basic, smallest, and non further separable phonetic-phonological entities of speech production while the *syllable* forms the smallest speech unit which can be articulated or produced in isolation and which always consists of one or more gestures.

The interface level between the cognitive-linguistic and the sensorimotor module of speech production is the *phonological form level*. While the phonological form is the output of lexical processing in the production pathway, the sensorimotor (or phonetic-articulatory) processing begins with structuring this phoneme sequence in optimal sized chunks which can easily be converted in articulatory patterns. In most sensorimotor models of speech production, it is assumed that these chunks are syllables. The main task of the sensorimotor part of the production model is to convert the phonological syllable state into a set of executable motor commands. Thus, a cognitive-linguistic planning stage including syllabification is followed by a sensorimotor planning-programming state which subsequently generates articulatory movement patterns.

Some neurologically based theories of speech production differentiate motor planning and motor programming (e.g., [2]). Here, it is assumed that a *core motor plan* comprises inverse internal models of movements which contain spatial information (place and manner of articulation) and

temporal information (inter-articulatory movement synchronization; [2], p.405). Because an inverse model transforms a desired sensory outcome directly into motor commands the related motor plan comprises motor *and* sensory information. The subsequently generated *motor program* in addition comprises muscle specific and articulator and muscle specific control information regarding muscle tone, movement direction, velocity, force, range, and mechanical stiffness of joints ([2], p. 406). While motor plans are specified at the level of the premotor cortex (i.e., the location of the speech sound map in the DIVA model, see [3]), motor programs define neural activations at the primary motor cortex which subsequently allows the direct execution of the speech item.

In this contribution we will focus on modeling motor programming. Because motor programming not only gives a full specification of gesture scores for each syllable (for a definition of gesture scores see [6-8]) but as well a quantitative specification of the muscle activation pattern of all muscles controlling all vocal tract articulators, it is necessary to specify the contribution of each articulator to each vocal tract movement unit (gesture) at the motor program level.

2 Architecture of the model

In our neurobiologically inspired quantitative and computer-implemented production model four main modules can be separated (cf. [9, 10]): (i) linguistic planning (from concepts via lemmas towards phonological forms including syllabification, cf., [1]), (ii) motor planning (establishing a *raw gesture score*), (iii) motor programming (establishing a *fully specified gesture score* plus single articulator contributions to each gesture and plus specification of temporal *muscle activation patterns* for execution), and (iv) execution (i.e., generation of articulatory movements and of the acoustic speech signal). While linguistic planning has been introduced in [9] for our modeling approach, we will describe below the sensorimotor part of our model.

2.1 Planning and programming: motor plans and motor programs

The *raw gesture score* (which results from motor planning and which alternatively can be called *motor plan*) identifies all gestures making up a syllable (e.g., vocalic tract-shaping gestures, consonantal constriction-forming gestures, velopharyngeal opening and closing gestures, glottal opening and closing gestures) in a distinctive qualitative way (see [6-8]) and it identifies the raw temporal coordination of gestures as prescribed by the ordering of segments within the sound sequence of the syllable. In the case of the syllable /pa/ (e.g., in Standard German) a labial closing gesture is temporally overlapping with a velopharyngeal closing and a glottal opening gesture for realizing the voiceless plosive /p/. This labial closing gesture is partly overlapping and partly preceding a vocalic tongue lowering gesture for producing the /a/ which in addition is temporally overlapping with a glottal closing gesture for realizing the phonation for the vowel. It can be shown that our approach for planning is mainly a conversion of phonological (or broad phonetic) information into distinctive articulatory information without giving a precise temporal specification of gesture activation time intervals as well as of muscle activation levels.

Thus, a motor plan or raw gesture score is defined as *phonological level specification* of all gestures building up a syllable. But in contrast to the phoneme string, which can be seen as a *perception-related phonological form*, the raw gesture score can be interpreted as a *production-related phonological form* (articulatory-phonological form). This form comprises (i) all gestures building up the segments within the syllable, (ii) the constriction type, constriction location (i.e., gesture target specification), the gesture executing end-articulator per gesture, and (iii) the raw temporal coordination of gestures with each other per syllable (cf., [6], p. 17ff). Because gesture targets are defined in the *tract variable space* (i.e., defining global or local vocal tract shapes; see [7, 8]) targets are defined in terms of articulation but they are closely related to the acoustic-auditory domain as well.

The motor plan is the basis for the calculation of the *motor program* also called *fully specified gesture score*. Thus, while motor plans or raw gesture scores specify gesture types and their raw temporal coordination, motor programs in addition specify the exact temporal appearance of all gestures, their exact target-reaching trajectories, as well as the contribution of all primary and secondary articulators to each gesture. Here, *primary articulators* are the constriction-forming end-articulators (e.g., lips in case of labial consonants like /p, b, m/, tongue tip in case of apical consonants like /t, d, n/, tongue dorsum in case of dorsal consonants like /k, g, N/; SAMPA notation is used here, see [11]) while *secondary articulators* support the gesture-induced movements of the primary articulators (e.g., lower

jaw as secondary articulator supports labial, apical, and dorsal constriction-forming gestures as well as vocalic tract-shaping gestures, while lower jaw and tongue dorsum as secondary articulators support apical constriction-forming gestures). After fully specifying the gesture score, our model is capable to calculate the neural activation patterns over time for all muscles of all articulators involved in the production of each syllable. This overall neuromuscular activation pattern is the output of the motor programming module or input specification for articulatory execution of a syllable.

The model controls an *articulatory-acoustic speech synthesizer* [12]. The input for this synthesizer is the neuromuscular activation pattern (muscle activation pattern). *Muscle groups* are defined here for specifying the *main movement directions* for each articulator. It will be shown that vocalic tract-shaping as well as consonantal constriction-forming gestures are controlled by complex muscle activation patterns which can be understood best if they are interpreted in the concept of co-contraction of different *functional muscle groups* each comprising sets of two or three muscles (see chapter 2.2 of this paper). Each functional muscle group is involved in moving articulators in *one* specific main movement direction (cf., [13]). Thus, an important feature of our articulator model is the separation of *functionally different movements (movement directions)* of articulators, which as well is a separation of consonantal versus vocalic movements and movement directions. Moreover, articulators can be divided each gesture in gesture-executing end-articulators (gesture-executing primary articulators) and gesture-executing secondary articulators as already introduced above. We will present preliminary quantitative rules for separating muscle activation between muscle groups, which are related to primary and secondary articulators for each type of gesture and thus for quantifying the contribution of each articulator to each gesture (see chapter 3 of this paper).

2.2 The functional articulator model: articulators and muscle groups

Tongue positions and movements are mainly controlled by *six extrinsic muscles,* controlling tongue dorsum movements – i.e., the genioglossus anterior part (GGa), middle part (GGm), posterior part (GGp), the hyoglossus (HG), and the styloglossus (SG) – and by *four intrinsic muscles*, mainly controlling the tongue tip movements – i.e., the superior longitudinalis (SL), the inferior longitudinalis (IL), the transversus (T), and the verticalis (V) [13, 14]. *Vocalic tongue shapes* (main part of the global vocal tract shapes) are mainly controlled by extrinsic tongue muscles: (i) high-front /i/-like position involves the activation of the GGp and, to a much lesser extent, of the SG; (ii) low /a/-like position involves the activation of the HG and of the GGa; (iii) high-back /u/-like position involves the activation of the SG, and, to a much lesser extent, of the GGp (see [14], p. 1589). *Consonantal tongue shapes* are controlled by extrinsic and intrinsic tongue muscles for the *tongue dorsum* (/k/-like elevation of tongue dorsum involves activation of the SG, the GGp, and, to a lesser extent, of the IL; [14], p. 1589), as well as for the *tongue tip.* (i) The /t/-like elevation of tongue tip involves activation of the SL and a T-SL combination (see [13], p. 863ff) or activation of a GGp-SL combination (see [13], p. 866). (ii) The tongue tip lowering involves activation of the V (see [13], p. 862) or activation of a GGa-IL combination (see [13], p. 865).

The muscles controlling the *low-high (opening-closing) movement direction of the lower jaw* can be grouped as an agonist-antagonist pair of muscle groups. jaw-openers are mainly the anterior belly of digastric and the geniohyoid. Jaw-closers are mainly the masseter and the medial pterygoid (see [15], p. 375 and [13], p. 858). A further important movement is that of the hyoid bone, controlling the height of the larynx (including the glottis) and thus controlling the position of the lower ending of the vocal tract and thus together with the tongue and the lips controlling overall length of the vocal tract. Here exists an agonist-antagonist pair of muscle groups which controls the *upward-downward movement direction of the hyoid bone*. The upward retracting movement ins mainly controlled by the posterior belly of digastric and by the stylohyoid. The downward depressing movement is mainly controlled by the sternohyoid, the thyrohyoid, and the omohyoid (see [15], p. 375).

An *upward movement of the velum* results from the contraction of the levator veli palatini while its relaxation leads to a downward movement of the velum [16]. Thus, the velum movement is mainly controlled by only one muscle (group) and not by an agonist-antagonist pair of muscle groups. But while an opening of the velopharyngeal port is produced easily by relaxation (deactivation) of the levator veli palatini, which not needs to be a strong opening, but simply a kind of a “leak” of the velopharyngeal port in case of nasals, the closure of the velopharyngeal port (upward movement of the velum) needs to be strong and tight in case of plosives and fricatives. In the case of a low activation

level or dysfunction of this muscle velopharyngeal closures may stay incomplete. This leads to a disordered production (to a nasalization) especially of plosives and fricatives because the needed air pressure cannot be built up in the oral cavity during the time interval of closure or constriction.

Lip closing is accomplished by activating the orbicularis oris superior (OOS), the depressor anguli oris (DAO), and the mentalis, while *mouth opening* mainly results from activating the depressor labii inferioris (DLI) and levator labii superioris (LLS). Lip opening if often accompanied by movemetns of secondary articulators like the lower jaw. *Lip protrusion* results from activation of the orbicularis oris inferior (OOI) and of the depressor labii inferior (DLI) while *lip spreading* mainly results from activation of the risorius [17, 18].

Thus, *pairs of muscle groups can be found for each of the main movement directions of all articulators* acting as agonist-antagonist pairs for each movement direction. Moreover, the separate grouping of muscles controlling tongue dorsum and tongue tip movements motivates that it is advantageous to separate consonantal and vocalic articulation not just on the higher level of motor planning (separation of vocalic and consonantal gestures, see [6-8]) but as well on the lower muscle-related level. From the summary of vocal tract muscle functionality given above it can be concluded that vocalic as well as consonantal gestures are controlled by complex muscle activation patterns which can be understood best if they are interpreted in the concept of co-contraction of organized agonist-antagonist pairs of muscle groups. Each muscle group consists of one, two or even more than two muscles (cf. [13], p. 865ff). This allows to establish *functional model muscle groups* (Tab. 1) for controlling gestures and subsequently for controlling the *articulators* within our functional articulator model (Tab. 2; for a detailed description of the model see [12]).

Table 1 - List of functional model muscle groups for controlling the model articulators (up-down movements of the hyoid are considered in addition for vocalic articulation).

name of agonist-antagonist model muscle groups	**controlled articulators**	**controlled movement direction (and movement type)**
jaw raising-lowering muscle groups	jaw	low-high
velum raising-lowering muscle groups	velum	low-high
tongue dorsum raising-lowering muscle groups	tongue dorsum	vocalic-raised (consonantal)
tongue tip raising-lowering muscle groups	tongue tip	vocalic-raised (consonantal)
lips closing-opening muscle groups	lips	vocalic-closed (consonantal)
lips rounding-unrounded muscle groups	lips	unrounded-rounded
tongue dorsum vocalic high-low muscle groups	tongue dorsum	low-high (vocalic)
tongue dorsum vocalic front-back muscle groups	tongue dorsum	back-front (vocalic)
hyoid raising-lowering muscle groups	hyoid	low-high

Thus, an important feature of our *functional articulator model* is the separation of *functionally different movements of articulators* (consonantal or vocalic) as well as the separation of gesture-executing articulators in gesture-executing end-articulators (*primary articulators*: lips, tongue tip, tongue dorsum, velum; see Tab. 2) and *secondary articulators*. The primary articulators establish vocal tract constrictions, and the secondary articulators are those on which the primary articulators depend and which assist in realizing a concrete gesture (see Tab. 2). An articulator can be a primary or secondary articulator, depending on the gesture (or task) under execution. For example, the tongue dorsum is a primary articulator for vocalic tract-shaping and consonantal dorsal constriction-forming gestures, while tongue dorsum is a secondary articulator in the case of consonantal apical constriction-forming gestures.

Table 2 - List of gesture-executing end-articulators and its related secondary articulators, also involved in gesture execution.

gesture-executing end-articulator (primary articulator)	**secondary articulators**
lips	jaw
tongue tip	jaw, tongue dorsum
tongue dorsum (dorsal consonantal gestures)	jaw
tongue dorsum (vocalic gestures)	jaw, hyoid
velum	-

Tab. 1 and Tab. 2 illustrate, which functional model muscle groups are involved in the execution of different types of gestures regarding to the primary and secondary articulators. Moreover, movement type, primary articulator, secondary articulators, and the movement directions are defined for each gesture (see Tab. 3 and Fig. 1). This subsequently defines the number of muscle groups which are

involved in the execution of each gesture. A list of gestures for Standard German is given in [6]: vocalic gestures are labeled here as vocalic tract-shaping gestures, consonantal gestures are labeled here as constriction-forming gestures, forming a labial, an apical or a dorsal constriction; velopharyngeal and glottal opening or closing gestures separate nasal from oral sounds and voiced from voiceless sounds.

Table 3 - Articulators and movement directions involved in the execution of different types of gestures. Extremal articulator positions are shown in Fig. 1 for each gesture listed in this table.

name of gesture type	**articulators involved (primary, secondary)**	**movement directions and types**	**example in Fig. 1**
vocalic tract shaping gesture (tongue dorsum part)	tongue dorsum, jaw	low-high (vocalic), low-high	a
	tongue dorsum, hyoid	back-front (vocalic), low-high	b
labial closing gesture	lips, jaw	vocalic-closed (consonantal), low-high	d
apical closing gesture	tongue tip, tongue dorsum, jaw	vocalic-raised (consonantal), vocalic-high, vocalic-front, low-high	e
dorsal closing gesture	tongue dorsum, jaw	vocalic-raised (consonantal), low-high	f
velopharyngeal opening or closing gesture	velum	low-high	c

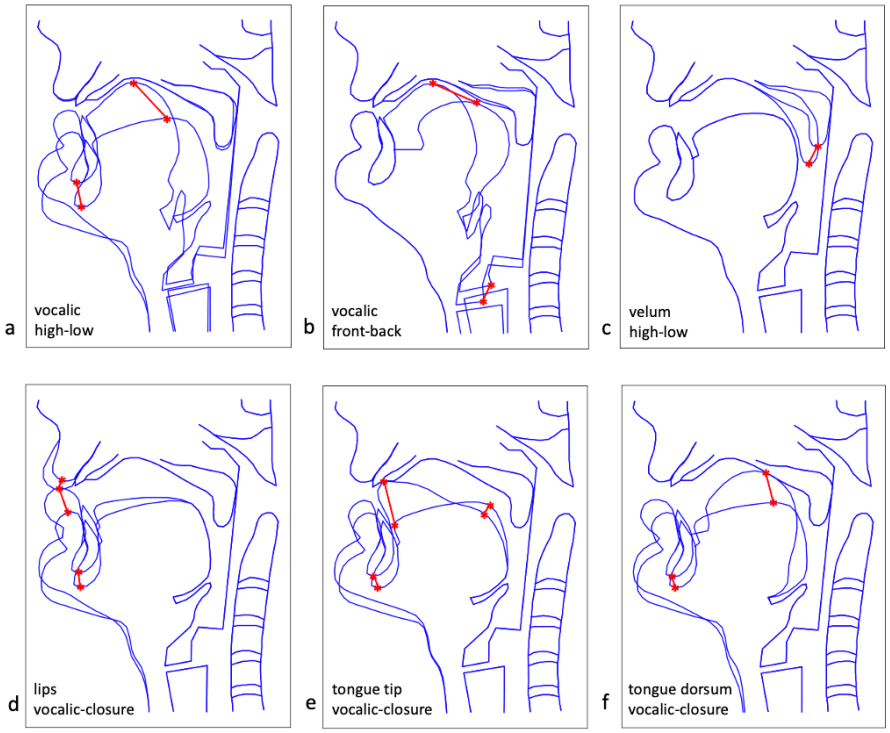

Figure 1 - Midsagittal views of vocal tract shapes for extremal articulator positions in case of different gestures defined in Tab. 3. The vocal tract shapes are generated using our articulator model [10]. The term "vocalic" indicates the current vocal tract opening at the lips, tongue tip or tongue dorsum before the closing gesture starts (Fig. 1c, 1d, 1e).

3 Preliminary results: a simulation example

The gesture activation for some consonantal gestures (vocal tract constriction-forming gestures) as well as for vocalic gestures (vocal tract-shaping gestures) is displayed for a sequence of three syllables of Standard German in Fig. 2 (nonsense word). The activation level as function of time specifies the degree of gesture realization. This specifies the absolute location of all *primary* articulators for all gestures and subsequently the vocal tract shapes over time for the whole syllable sequence (Fig. 3). The relative (as well as absolute) location of *all* articulators including all secondary articulatory can be calculated in a further step following the calculation of the trade-off factor for primary and secondary articulator contribution for vocalic and consonantal gestures (Fig. 4). This needs an approximation of the degree with which secondary articulators within a gesture participate in the target-reaching process for each type of gesture.

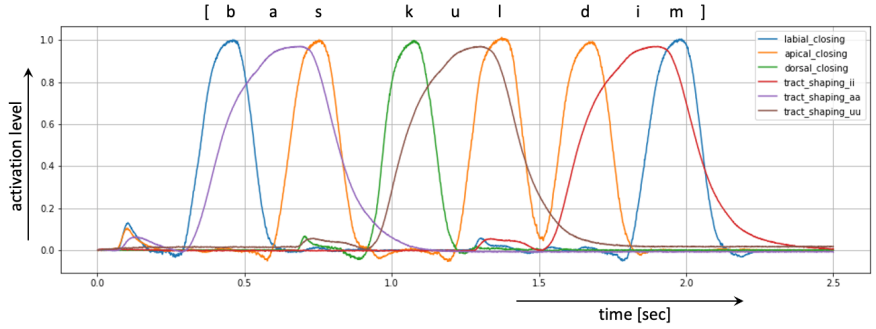

Figure 2 - Activation levels of the vocalic and consonantal gestures for a realization of the syllable sequence /bas.kul.dim/. The syllables are uttered as a sequence with short pauses. (Phonetic realization in transcription brackets []).

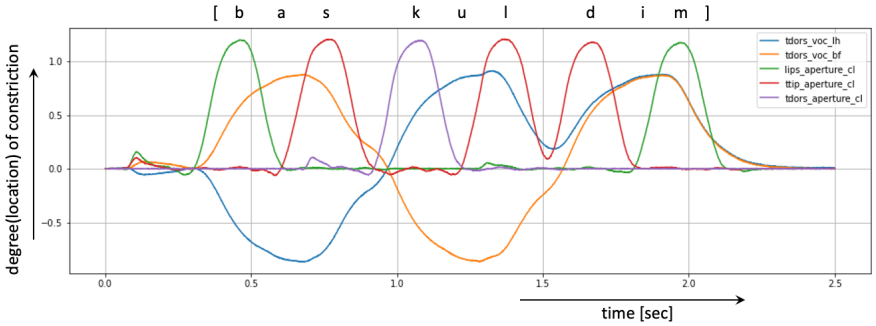

Figure 3 - Location of primary articulators as function of time for the specification of vocalic vocal tract shapes (degree and location of constriction) and consonantal vocal tract constrictions (degree of constriction) for a realization of the syllable sequence /bas.kul.dim/. Abbreviations within legend of Fig. 3: ttip = tongue tip; tdors = tongue dorsum; cl = consonantal closing gesture; voc = vocalic gesture; lh = low-high; bf = back-front.

The three simulated syllables /bas.kul.dim/ (nonsense word in Standard German) comprise labial, apical, and dorsal consonantal closing gestures as well as three vocalic gestures (Fig. 2). The gesture activation levels can be converted in target-directed trajectories for specific (tract-)variables which define the displacement of the primary articulators for each gesture in the tract variable space (Fig.3). The target is specified as location and degree of constriction in case of consonantal closing gestures (degree >= 1.0 represents a closure or constriction; degree > 1.0 represent the strength of closure or constriction; the definition of location or place as well as of manner of articulation is done by further gesture parameters not mentioned here) and the target is the vocal tract shape in case of vocalic gestures (the vocal tract shape is parameterized by values for degree tdors_voc_lh which range from -1 (low) to +1 (high) and by values for location tdors_voc_bf which range from -1 (back) to +1 (front); the rounded-unrounded parameterization is not visualized in Fig. 3). All consonantal gestures reach values for degree > 1 indicating that the consonantal constriction or closure is really reached (lips_..., ttip_..., and tdors_aperture_cl in Fig. 3). The vocalic gestures reach values at about 0.8 which represent the low-front target for the /a/-realization, the high-back target for the /u/-realization and the high-front target for the /i/-realization (tdors_voc_lh, and …_bf in Fig. 3).

These displacement trajectories or tract-variable trajectories (Fig. 3) already define the absolute positions of the primary articulators in the case of the consonantal gestures (end-articulators: lips, tongue tip and tongue dorsum) as well as the absolute position for the whole tongue body in case of the vocalic gestures (low-high and front-back). But in case of all consonantal and vocalic gesture the movement of a secondary articulator like that of the lower jaw needs to be specified in addition. In case of all vocalic and all consonantal gestures the location of the hyoid and larynx needs to be specified as well.

In our model the trade-off factor between contribution of lower jaw and tongue dorsum is set to 0.4 for the how-high movement direction in case of vocalic gestures (cf. Fig. 1a) based on static midsagittal MRI scans (long vowels, schwa, and consonants of Standard German hold at their maximum constriction of Standard German, see [12]). Thus, in case of vocalic gestures the lower jaw contributes 40% and the tongue dorsum contributes 60% to the vocal tract shaping (raw estimate). This leads to a lower mean value of jaw position during the realization of the syllable /bas/ in comparison to the realizations of the syllables /kul/ and /dim/ (see Fig. 4).

The trade-off factor between the contribution of lower jaw and the primary articulators in case of consonantal closing gestures is set to 0.5 for labial and to 0.3 for apical and dorsal consonantal

gestures. Thus, in case of these consonantal closing gestures the jaw contributes 30%-50% and the primary articulator contributes 50%-70% the closing gesture (raw estimate). This contribution results in an upward-downward movement of the lower jaw during the time intervals of the consonantal closing gestures for all three syllables (strongest for labial closing gestures, see Fig. 4). For apical consonantal gestures the 70% contribution of tongue dorsum and tongue tip is split in a half-half contribution (raw estimate). In case of a low vowel context the jaw contribution seems to be low (Fig. 4, realization of syllable /bas/) but it should be kept in mind that the consonantal lower jaw movement here starts from negative values (low jaw position because of low vowel context). The maximum relative displacement of the end-articulator is higher in case of consonantal gestures in a low vowel context than in a high vowel context. This reflects the fact that these relative displacement values are defined with respect to the current position of the lower jaw. But the lower jaw values m_jaw_lh *plus* relative displacement values for the articulators involved in consonantal closing gestures (m_lips_nc, m_ttip_nc, or m_tdors_nc) lead to absolute displacements > 1 for primary articulators and thus guarantee the production of a consonantal closure or constriction.

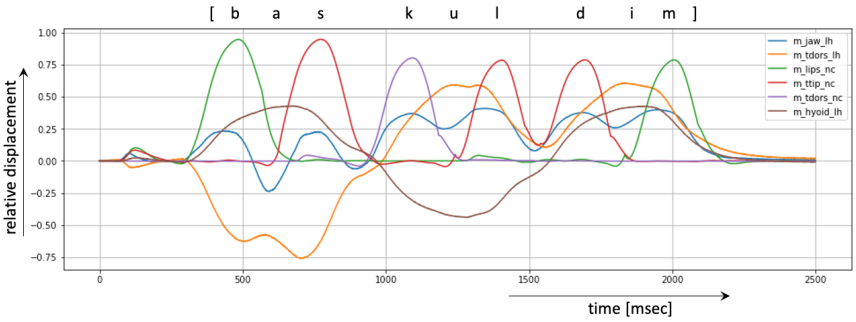

Figure 4 - Relative articulator displacement for the vertical movement direction of lips, tongue tip, tongue dorsum, lower jaw, and hyoid bone for the uttered syllable sequence /bas.kul.dim/ (trade-off factor is 0.5 here for all consonantal gestures; tongue dorsum-tip trade-off for apical gestures is not included).

It should be mentioned that the displacement of the hyoid bone in our model is mainly a function of the vocalic front-back parameter. Thus, this gesture parameter controls one main movement direction of the tongue body as well as that of the hyoid bone (see m_hyoid_lh in Fig. 4 in comparison to tdors_voc_bf in Fig. 3; and see Fig. 1b).

The displacement values for all articulators relative to each other can be transferred directly into neural activation levels of agonist-antagonist pairs of muscle groups controlling a movement direction of an articulator. (Thus, all relative displacement variables begin with "m_" like "muscle", see legend of Fig. 4). The displacement values shown in Fig. 4 are relative articulator displacements, representing articulator displacements between -1 (maximum negative displacement; i.e., low) and +1 (maximum positive displacement; i.e., high). A maximum relative displacement value (+1) for an articulator coincides with a maximum(minimum) activation level of its agonist(antagonist) model muscle group while a minimum relative displacement value (-1) for an articulator coincides with a minimum(maximum) activation level of its agonist(antagonist) model muscle group.

Relative displacement values are always positive for the consonantal closing gestures. These gestures start from different neutral positions which are defined by the vocalic context (e.g., high neutral position and thus only a short distance towards the closing target position in case of high vowels; low neutral position and thus a long distance towards the closing target position in case of low vowels). This feature is reflected in the naming of these parameters with the ending "_nc" (see legend of Fig. 4) which means "neutral-to-closure".

4 Discussion and conclusions

The neurobiologically based quantitative and computer-implemented model of speech production introduced here allows us to elucidate specific sub-processes of speech production like motor planning, motor programming and execution. Moreover, the model gives insights how all different muscles involved in articulation are organized functionally in speech articulation. Last but not least our approach highlights the idea of vocal tract different types of gestures. It can be concluded that *the concept of speech movement units* (or *gestures* [6-8]) is very helpful for getting a better understanding of the complex spatio-temporal processes taking place in speech articulation.

The planning-programming dichotomy has been discussed and illustrated for our speech production model. While acoustic-auditory as well as articulatory tract-variable targets can be set already on the planning level, the exact temporal specification of gesture on- and offsets, the exact contribution of all articulators to each gesture, and the detailed muscle activation patterns controlling all articulators over time are specified on the programming level. Thus, a motor plan needs to be programmed before it can be executed. Because of the complexity of the muscle system, it is advantageous to define a *small set of agonist-antagonist model muscle groups* for each main movement direction and movement type (vocalic vs. consonantal) of each articulator. In this case muscle activation can be deduced from relative articulator displacements in a basic production model as introduced here.

List of References

[1] LEVELT, W.J.M., ROELOFS, A. AND A.S. MEYER (1999) *A theory of lexical access in speech production*. Behavioral and Brain Sciences 22, 1-75.

[2] VAN DER MERWE, A. (2021) *New perspectives on speech motor planning and programming in the context of the four-level model and its implications for understanding the pathophysiology underlying apraxia of speech and other motor speech disorders*. Aphasiology 35, 397-423.

[3] GUENTHER, F.H., GHOSH, S.S. AND J.A. TOURVILLE (2006) *Neural modeling and imaging of the cortical interactions underlying syllable production*. Brain and Language 96, 280-301.

[4] BOHLAND, J.W., BULLOCK, D. AND F.H. GUENTHER (2010) *Neural Representations and Mechanisms for the Performance of Simple Speech Sequences.* Journal of Cognitive Neuroscience 22, 1504-1529.

[5] MILLER, H.E. AND F. H. GUENTHER (2021) *Modelling speech motor programming and apraxia of speech in the DIVA/GODIVA neurocomputational framework*. Aphasiology 35, 424-441.

[6] KRÖGER, B.J. AND T. BEKOLAY (2019) *Neural Modeling of Speech Processing and Speech Learning. An Introduction.* Springer Nature, Cham, Switzerland.

[7] BROWMAN, C.P. AND L. GOLDSTEIN (1992) Articulatory phonology: An overview. Phonetica 49, 155-180.

[8] GOLDSTEIN, L., BYRD, D. AND E. SALTZMAN (2006) *The role of vocal tract gestural action units in understanding the evolution of phonology*. In: M. A. Arbib (ed.) *Action to language via the mirror neuron system* (Cambridge, UK: Cambridge University Press), pp. 215-249.

[9] KRÖGER, B.J., STILLE, C.M., BLOUW, P., BEKOLAY, T. AND T.C. STEWART (2020) *Hierarchical Sequencing and Feedforward and Feedback Control Mechanisms in Speech Production: A Preliminary Approach for Modeling Normal and Disordered Speech*. Frontiers in Computational Neuroscience 14:573554.

[10] KRÖGER, B.J. (2021) *Modeling dysfunctions in the coordination of voice and supraglottal articulation in neurogenic speech disorders*. In: C. Manfredi (ed.) Models and Analysis of Vocal Emissions for Biomedical Applications. (Firenze, Italy: Firenze University Press), pp. 79-82.

[11] SAMPA – computer-readable phonetic alphabet (also: Speech Assessment Methods Phonetic Alphabet). https://www.phon.ucl.ac.uk/home/sampa/, accessed at December 16th, 2021.

[12] KRÖGER, B.J., BEKOLAY, T. AND C. ELIASMITH (2014) *Modeling speech production using the Neural Engineering Framework*. Proceedings of CogInfoCom 2014 (Vetri sul Mare, Italy), pp. 203-208. IEEE Xplore Digital Library, doi=10.1109/CogInfoCom.2014 .7020446.

[13] DANG J. AND K. HONDA (2004) *Construction and control of a physiological articulatory model*. Journal of the Acoustical Society of America 115, 853-870.

[14] PERRIER, P., PAYAN, Y., ZANDIPOUR, M. AND J. PERKELL (2003) *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.

[15] LABOISSIÈRE, R., OSTRY, D.J. AND A.G. FELDMAN (1996) *The control of multi-muscle systems: human jaw and hyoid movements*. Biological Cybernetics 74, 373-384.

[16] PELLAND, C.M., FENG, X., BOROWITZ, K.C., MEYER, C.H. AND S.S. BLEMKER (2019) *A dynamic magnetic resonance imaging–based method to examine in vivo levator veli palatini muscle function during speech*. Journal of Speech Language and Hearing Research 62, 2713-2722.

[17] HONDA, K., KURITA, T., KAKITA, Y. AND S. MAEDA (1995) *Physiology of the lips and modeling of lip gestures*. Journal of Phonetics 23, 243-254.

[18] KING, S.A., PARENT, R.E. AND B.L. OLSAFSKY (2000) *An anatomically-based 3D parametric lip model to support facial animation and synchronized speech*. Proceedings of the IFIP TC5/WG5.10 DEFORM'2000 Workshop and AVATARS'2000 Workshop on Deformable Avatars (Deventer, NL: Kluwer), pp. 12-23.